

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C07K 14/705, A61K 38/17, C12N 5/10, 15/12</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 99/11668</b> <b>(43) International Publication Date:</b> 11 March 1999 (11.03.99)
<b>(21) International Application Number:</b> PCT/US97/15695 <b>(22) International Filing Date:</b> 3 September 1997 (03.09.97)  <b>(71) Applicant (for all designated States except US):</b> THE UNIVERSITY OF MISSISSIPPI MEDICAL CENTER [US/US]; 2500 North State Street, Jackson, MS 39216-4505 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> PILETZ, John, E. [US/US]; 14 Peachtree Lane, Madison, MS 39110 (US). IVANOV, Tina, R. [CA/US]; 2590 N. State Street, Jackson, MS 39216 (US).  <b>(74) Agent:</b> CHEEK, Warren, M., Jr.; Wenderoth, Lind & Ponack, Suite 800, 2033 "K" Street, N.W., Washington, DC 20006 (US).		<b>(81) Designated States:</b> AU, CA, HU, IL, JP, NZ, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> DNA MOLECULES ENCODING IMIDAZOLINE RECEPTIVE POLYPEPTIDES AND POLYPEPTIDES ENCODED THEREBY		
<b>(57) Abstract</b>  A genomic DNA encoding a human imidazoline receptor is described. cDNAs encoding the receptor and fragments thereof are also provided. An amino acid sequence predicted to be 120,000 MW for nearly the entire protein is identified, as well as a middle fragment believed to contain the imidazoline binding site of the receptor. The protein is highly unique in its sequence and may represent the first in a novel family of receptor proteins. Methods of cloning the cDNA and expressing the imidazoline receptor in a host cell are described. Methods of preparing antibodies against the transfected protein are also described. Also, a screening method for identifying additional subtypes of this receptor are identified. Also, screening methods for identifying drugs that interact with the imidazoline receptor are described.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

**DNA MOLECULES ENCODING IMIDAZOLINE RECEPTIVE POLYPEPTIDES  
AND POLYPEPTIDES ENCODED THEREBY**

BACKGROUND OF THE INVENTION

1. Field of the Invention

5           The present invention is directed to DNA molecules encoding imidazoline receptive polypeptides, preferably encoding human imidazoline receptive polypeptides, that can be used as an imidazoline receptor (abbreviated IR). In addition, transcript(s) and protein sequences are predicted from the DNA  
10 clones. The invention is also directed to a genomic DNA clone designated as JEP-1A. The cDNA clones according to the invention comprise cDNA homologous to portion(s) of this genomic clone; including 5A-1 cDNA, cloned by the inventors that established the open-reading frame for translation of  
15 mRNA from the gene, and established the immunoreactive properties of its polypeptide sequence in an expression systems. Also, the invention relates to cDNA clone EST04033, which is another clone identified to contain cDNA sequences from the JEP-1A gene, and of which the 5A-1 is a part, that  
20 encodes an active fragment of the IR polypeptide in transfection assays, and the protein sequences thereof. The invention also relates to methods for producing such genomic and cDNA clones, methods for expressing the IR protein and fragments, and uses thereof.

25   2. Description of Related Art

It is believed that brainstem imidazoline receptors possess binding site(s) for therapeutically relevant imidazoline compounds, such as clonidine and idazoxan. These drugs represent the first generation of ligands discovered for the binding site(s) of imidazoline receptors. However, clonidine and idazoxan were developed based on their high affinity for  $\alpha_2$ -adrenergic receptors. Second generation ligands, such as moxonidine, possess somewhat improved selectivity for IR over  $\alpha_2$ -adrenergic receptors, but more selective compounds for IR are needed.

An imidazoline receptor clone is of particular interest because of its potential utility in identifying novel pharmaceutical agents having greater potency and/or more selectivity than currently available ligands have for imidazoline receptors. Recent technological advances permit pharmaceutical companies to use combinatorial chemistry techniques to rapidly screen a cloned receptor for ligands (drugs) binding thereto. Thus, a cloned imidazoline receptor would be of significant value to a drug discovery program.

Until now, the molecular nature of imidazoline receptors remains unknown. For instance, no amino acid sequence data for a novel IR, e.g., by N-terminal sequencing, has been reported. Three different techniques have been described in the literature by three different laboratories to visualize imidazoline-selective binding proteins (imidazoline receptor candidates) using gel electrophoresis. Some important consistencies have emerged from these results despite the diversity of the techniques employed. On the other hand,

multiple protein bands have been identified, which suggests heterogeneity amongst imidazoline receptors. These reports are discussed below.

Some of the abbreviations used hereinbelow, have the following meanings:

	$\alpha_2$ AR	Alpha-2 adrenoceptor
	BAC	Bovine adrenal chromaffin
	ECL	Enhanced chemiluminescence (protein detection procedure)
10	EST	Expressed Sequence Tag (a one-pass cDNA documentation without identification)
	I-site	Any imidazoline-receptive binding site (e.g., encoded on IR)
15	IR <sub>1</sub>	Imidazoline receptor subtype <sub>1</sub>
	IR-Ab	Imidazoline receptor antibody
	I <sub>2</sub> Site	Imidazoline binding subtype <sub>2</sub>
	kDa	Kilodaltons (molecular size)
	MAO	monoamine oxidase
	MW	molecular weight
20	NRL	European abbreviation for RVLM (see below)
	PC-12	Phaeochromocytoma-12 cells
	<sup>125</sup> I PIC	[ <sup>125</sup> I]p-iodoclonidine
	PKC	Protein Kinase C
	RVLM	Rostral Ventrolateral Medulla in brainstem
25	SDS	sodium dodecyl sulfate gel electrophoresis

Reis et al. [Wang et al., Mol. Pharm., 42: 792-801 (1992); Wang et al., Mol. Pharm., 43: 509-515 (1993)] were the first to characterize an imidazoline-selective binding protein and to demonstrate it as having MW = 70 kDa. This was accomplished using bovine cells (BAC), which lack an  $\alpha_2$ AR [Powis & Baker, Mol. Pharm., 29:134-141 (1986)]. The 70 kDa imidazoline-selective protein in those studies had high affinities for both idazoxan and p-aminoclonidine affinity chromatography columns and was eluted by another imidazoline compound (phentolamine). Unfortunately, those investigators failed to isolate sufficient 70 kDa protein to determine its other biochemical properties. To date, no one has reported the complete purification of an imidazoline receptor protein.

Likewise, no amino acid sequences have been reported for IR.

Their 70 kDa protein was used by Reis and co-workers to raise "I-site binding antiserum", designated herein as Reis antiserum. The term "I-site" refers to the imidazoline binding site, presumably defined within the imidazoline receptor protein. Reis antiserum was prepared by injecting the purified protein into rabbits [Wang et al, 1992]. The first immunization was done subcutaneously with the protein antigen (10  $\mu$ g) emulsified in an equal volume of complete Freund's adjuvant, and the next three booster shots were given at 15-day intervals with incomplete Freund's adjuvant. The polyclonal antiserum has been mostly characterized by immunoblotting, but radioimmunoassays (RIA) and/or conjugated assay procedures, i.e., ELISA assays, are also conceivable [see "Radioimmunoassay of Gut Regulatory Peptides: Methods in Laboratory Medicine," Vol. 2, chapters 1 and 2, Praeger Scientific Press, 1982].

The present inventors and others [Escriba et al., Neurosci. Lett. 178: 81-84 (1994)] have characterized the Reis antiserum in several respects. For instance, the present inventors have discovered that human platelet immunoreactivity with Reis antiserum is mainly confined to a single protein band of MW  $\approx$  33 kDa, although a trace band at  $\approx$  85 kDa was also observed. The  $\approx$  33 and  $\approx$  85 kDa bands were enriched in plasma membrane fractions as expected for an imidazoline receptor. Furthermore, the intensity of the  $\approx$  33 kDa band was found to be positively correlated with non-adrenergic  $^{125}$ PIC Bmax values at platelet IR<sub>1</sub> sites in samples from the same

subjects, with an almost one-to-one slope factor. In addition, the nonadrenergic <sup>125</sup>PIC binding sites on platelets were discovered by the present inventors to have the same rank order of affinities as IR<sub>1</sub> binding sites in brainstem [Piletz and Sletten, J.Pharm. & Exper. Therap., 267: 1493-1502 (1993)]. The platelet  $\approx$  33 kDa band may also be a product of a larger protein, since in human megakaryoblastoma cells, which are capable of forming platelets in tissue cultures, an  $\approx$  85 kDa immunoreactive band was found to predominate.

Immunoreactivity with Reis antiserum does not appear to be directed against human  $\alpha_2$ AR and/or MAO A/B. This is a significant point because  $\alpha_2$ AR and MAO A/B have previously been cloned and also bind to imidazolines. The present inventors have obtained selective antibodies and recombinant preparations for  $\alpha_2$ AR and MAO A/B, and these proteins do not correspond to the  $\approx$  33, 70, or 85 kDa putative IR<sub>1</sub> bands. Thus, there is substantial evidence that, at least in human platelets, the Reis antiserum is IR<sub>1</sub> selective.

Another antiserum was raised by Drs. Dontenwill and Bousquet in France [Grenay et al., Europ. J. Pharmacol., 265: R1-R2 (1994); Grenay et al., Neurochem. Int., 25: 183-191 (1994); Bennai et al., Annals NY Acad. Sci., 763:140-148 (1995)] against polyclonal antibodies for idazoxan (designated Dontenwill antiserum). This anti-idiotypic antiserum inhibits <sup>3</sup>H-clonidine but not <sup>3</sup>H-rauwolscine ( $\alpha_2$ -selective) binding sites in the brainstem, suggesting it also interacts with IR<sub>1</sub> [Bennai et al., 1995]. As shown in Fig. 1, human RVLM (same as NRL) membrane fractions displayed bands of  $\approx$  41 and 44 kDa, as

detected by the present inventors using this anti-idiotypic antiserum.

The present inventors have found that the bands of MW  $\approx$  41 and 44 kDa detected by Dontenwill antiserum may be derived from an  $\approx$  85 kDa precursor protein, similar to that occurring in platelet precursor cells. An 85 kDa immunoreactive protein is obtained in fresh rat brain membranes only when a cocktail of 11 protease inhibitors is used. Also, as shown in Fig. 1, it is found that Reis antiserum detects the  $\approx$  41 and 44 kDa bands in human brain when fewer protease inhibitors are used. Additionally, the Dontenwill antiserum weakly detects a platelet  $\approx$  33 kDa band. Thus, the present inventors have hypothesized that the  $\approx$  41 and 44 kDa immunoreactive proteins may be alternative breakdown products of an  $\approx$  85 kDa protein, as opposed to the platelet  $\approx$  33 kDa breakdown product.

In summary, the main conclusion from the above results is that, despite vastly different origins, the Reis and Dontenwill antisera both detect identical bands in human platelets, RVLM, and hippocampus.

Using yet another technique, a photoaffinity imidazoline ligand,  $^{125}$ AZIPI, has also been developed to preferentially label  $I_2$ -imidazoline binding sites [Lanier et al., J.Biol.Chem., 268: 16047-16051 (1993)]. The  $^{125}$ AZIPI photoaffinity ligand was used to visualize  $\approx$  55 kDa and  $\approx$  61 kDa binding proteins from rat liver and brain. It is believed that the  $\approx$  61 kDa protein is probably MAO, in agreement with other findings [Tesson et al., J.Biol.Chem., 270: 9856-9861 (1995)] showing that MAO proteins bind certain imidazoline



compounds. The different molecular weights between these bands and those detected immunologically by the present inventors is one of many pieces of evidence that distinguishes IR<sub>1</sub> from I<sub>2</sub> sites.

5 To the inventors' knowledge and as described herein, we are first to clone the gene, cDNAs and fragments thereof encoding a protein with the immunological and ligand binding properties expected of an IR. On this basis, we are first to identify the nucleotide sequences of DNA molecules encoding an  
10 imidazoline receptor and active fragments thereof, and the first to determine the amino acid sequence of an imidazoline receptor and active fragments thereof. The polypeptides described herein are clearly distinct from  $\alpha_2$ AR or MAO A/B proteins.

15 SUMMARY OF THE INVENTION

The present invention involves various cDNA clones (ie., 5A-1 and EST04033) and a genomic clone (JEP-1A) which are directed to an isolated polypeptide(s) that is receptive to (bind to) imidazoline compound(s), and can be used to identify  
20 other compounds of interest. Currently available imidazoline compounds in this context are p-iodoclonidine and moxonidine. Initially, the inventors detected a polypeptide expressed by their cDNA clone (5A-1 isolated from a human hippocampus cDNA library) that immunoreacted with Reis antiserum and/or  
25 Dontenwill antiserum. The DNA sequence of the 5A-1 clone is encapsulated within a portion of the other clones (EST04033 and JEP-1A genomic clone).

In one aspect of the invention, a polypeptide includes a 651 amino acid sequence as shown in SEQ ID No. 5. This polypeptide is predicted from non-plasmid cDNA in EST04033; a clone which the inventors showed possesses sequences inclusive of 5A-1. Furthermore, transfection of EST04033 into COS cells yielded imidazoline receptivity by radioligand binding assays (described in detail later). Other imidazoline receptive proteins homologous to this polypeptide are also contemplated. Such polypeptide(s) generally have a molecular weight of about 50 to 80 kDa. More particularly, one can have a molecular weight of about 70 kDa.

In another aspect of this invention, a polypeptide includes a 390 amino acid sequence as shown in SEQ ID No. 6. This represents the polypeptide predicted from the non-plasmid DNA of the original 5A-1 clone. Such a polypeptide generally has a molecular weight of about 35 to 50 kDa. More particularly, it can have a molecular weight of about 43 kDa.

DNA molecules encoding aforementioned imidazoline-receptive polypeptide(s) are also contemplated. Such a DNA molecule, e.g., a cDNA derived from mRNA, can contain a nucleotide sequence encoding the 651 amino acid sequence shown in SEQ ID No. 5. Thus, a DNA molecule containing the 1954 base pairs (b.p.) (1954 b.p. encodes 651 amino acids) nucleotide sequence shown in SEQ ID No. 2 is contemplated. This represents the coding sequence for the polypeptide predicted by EST04033 transfections. In another embodiment, a DNA molecule includes the longer nucleotide sequence shown in SEQ ID No. 3. This represents the cDNA predicted to have been

translated + not predicted to have been translated in transfections experiments of EST04033.

In another embodiment of the invention, a DNA molecule contains a nucleic acid sequence encoding the amino acid  
5 sequence shown in SEQ ID No. 6. In another aspect, it can include the

1171 b.p. nucleic acid sequence shown in SEQ ID No. 4. The 1171 b.p. nucleic acid sequence shown in SEQ ID No. 4 is the 5A-1 non-plasmid DNA.

10 The nucleic acid sequence of the genomic clone encoding the imidazoline receptor is further shown in SEQ ID No. 21. The nucleic acid and amino acid sequence of the predicted transcript (ie., cDNA) can be predicted from the description hereinbelow. The polypeptide encoded by the genomic DNA is  
15 shown in SEQ ID No. 22.

Sequence similarity with the sequences indicated in SEQ ID protocols of the attached Sequence Listing is defined in connection with the present invention as a very close structural relationship of the relevant sequences with the  
20 sequences indicated in the respective SEQ ID protocols. To determine the sequence similarity, in each case the structurally mutually corresponding sections of the sequence of the SEQ ID protocol and of the sequence to be compared therewith are superimposed in such a way that the structural  
25 correspondence between the sequences is a maximum, account being taken of differences caused by deletion or insertion of individual sequence members (DNA-codon or amino acid respectively), and being compensated by appropriate shifts in

sections of the sequences. The sequence similarity in % results from the number of sequence members which now correspond to one another in the sequences ("homologous positions") relative to the total number of members contained in the sequences of the SEQ ID protocols. Differences in the sequences may be caused by variation, insertion or deletion of sequence members. Additionally in DNA sequences, different DNA-codons encoding for the same amino acid are considered identical in the context of the present invention. For amino acid sequences, conservative amino acid substitutions encoded by their corresponding DNA-codons, as well as naturally occurring homologs of the sequences, are considered within the context of sequence similarity.

DNA molecules of substantial homology ( $\geq 75\%$ ) are an implicit aspect of this sort of invention. As will be discussed later, the inventors have already identified two possible splice variants in the amino acid coding sequence. In addition, artificially mutated receptor cDNA molecules can be routinely constructed by methods such as site-directed polymerase chain reaction-mediated mutagenesis [Nelson and Long, Anal. Biochem. 180: 147-151 (1989)]. It is commonly appreciated that highly homologous mutants frequently mimic their natural receptor. A study by Kjelsberg et al. [J. Biol. Chem. 267: 1430-1433 (1992)] showed that all 20 amino acid substitutions produce an active receptor at a single site in the  $\alpha_{1b}$ -adrenergic receptor. RNA molecules of  $\geq 75\%$  complementarity to an instant DNA molecule, e.g., an mRNA molecule (sense) or a complementary cRNA molecule (antisense),

are a further aspect of the invention.

A further aspect of the invention is for a recombinant vector, as well as a host cell transfected with the recombinant vector, wherein the recombinant vector contains at least one of the nucleotide sequences shown in SEQ ID Nos. 1-4, or sequences predicted by the genomic clone, or nucleotide sequences  $\geq 75\%$  homologous thereto.

A method of producing an imidazoline receptor protein is another aspect of the invention. Such a method entails transfecting a host cell with an aforementioned vector, and culturing the transfected host cell in a culture medium to generate the imidazoline receptor.

A method for producing homologous imidazoline receptor proteins, and the proteins produced thereby, are also considered an aspect of this invention.

A significant further aspect of the invention is a method of screening for a ligand that binds to an imidazoline receptor. Such a method can comprise culturing an above-mentioned transfected cell in a culture medium to express imidazoline receptor proteins, followed by contacting the proteins with a labelled ligand for the imidazoline receptor under conditions effective to bind the labelled ligand thereto. The imidazoline receptor proteins can then be contacted with a candidate ligand, and any displacement of the labelled ligand from the proteins can be detected. Displacement of labelled ligand signifies that the candidate ligand is a ligand for the imidazoline receptor. These steps could be performed on intact host cells, or on proteins

isolated from the cell membranes of the host cells.

The invention will now be described in more detail with reference to specific examples.

#### BRIEF DESCRIPTION OF THE DRAWINGS

5           Fig. 1 depicts a comparison of Reis antiserum (lane 1, 1:2000 dilution) and Dontenwill antiserum (lane 2, 1:5000 dilution) immunoreactivities for human NRL (same as RVLM) and hippocampus, as discussed in Example 1.

10           Fig. 2 depicts a comparison of Reis antiserum (1:15,000 dilution) and Dontenwill antiserum (1:20,000 dilution) immunoreactivities for plaques isolated from the human hippocampal cDNA library used in cloning as discussed in Example 2. The plaques contain the initial clone, designated herein as 5A-1, in a third stage of purification.

15           Fig. 3 depicts the restriction map of the EST04033 cDNA clone.

20           Fig. 4 depicts a competitive binding assay between  $^{125}\text{I}$ -labelled p-iodoclonidine (PIC) and various ligands for the imidazoline receptor on membranes expressed in COS cells transfected with the EST04033 cDNA clone, as discussed in Example 4.

          Fig. 5 depicts the prediction of introns and exons of the genomic clone (as analyzed by the GENESCAN program and verified by the available CDNAS).

25           Fig. 6 depicts the distribution of MRNA homologous to our CDNA in human adult tissues (bar graph) and the two species of MRNA (6 and 9.5 kb).

DETAILED DESCRIPTION OF THE INVENTION

The present invention is concerned with multiple aspects of an imidazoline receptor protein, and DNA molecules encoding the same, and fragments thereof, which have now been  
5 discovered.

First, a polypeptide having imidazoline binding activity has been identified, which contains the putative active site for binding, as discussed hereinafter. Although polypeptide(s) described herein has a binding affinity for an  
10 imidazoline compound, it may also have an enzymatic activity, such as do catalytic antibodies and ribozymes. In fact, one such domain within our protein predicts a cytochrome p450 activity (described later).

Exemplary "binding" polypeptides are those containing  
15 either of the amino acid sequences shown in SEQ ID Nos. 5 or 6 (with the amino acid sequence predicted by EST04033 given in SEQ ID No. 5). Functionally equivalent polypeptides are also contemplated, such as those having a high degree of homology with such aforementioned polypeptides, particularly when they  
20 contain the Glu-Asp-rich region described hereinafter which we believe may define an active imidazoline binding site.

A polypeptide of the invention can be formed by direct chemical synthesis on a solid support using the carbodiimide method [R. Merrifield, JACS, 85: 2143 (1963)]. Alternatively,  
25 and preferably, an instant polypeptide can be produced by a recombinant DNA technique as described herein and elsewhere [e.g., U.S. Patent No. 4,740,470 (issued to Cohen and Boyer), the disclosure of which is incorporated herein by reference],

followed by culturing transformants in a nutrient broth.

Second, a DNA molecule of the present invention encodes  
aforementioned polypeptide. Thus, any of the degenerate set  
of codons encoding an instant polypeptide is contemplated. A  
5 particularly preferred coding sequence is the 1954 b.p.  
sequence set forth in SEQ ID No. 2, which has now been  
discovered to be a nucleotide sequence that encodes a  
polypeptide capable of binding imidazoline compound(s). In  
another embodiment, a DNA molecule includes the 3318 b.p.  
10 nucleotide sequence shown in SEQ ID No. 3. This latter  
sequence is the entire EST04033 insert. It includes the  
nucleotide sequence of SEQ ID No. 2 which was predicted to  
have been translated into protein in the transfection  
experiments.

15 In another embodiment of the invention, a DNA molecule  
contains a nucleic acid sequence encoding the amino acid  
sequence (390 residues) shown in SEQ ID No. 6. This amino  
acid sequence corresponds to that derived from direct  
sequencing of the 5A-1 clone and represents a fragment of the  
20 native protein. The 5A-1 DNA molecule is defined by the 1171  
b.p. nucleic acid sequence shown in SEQ ID No. 4.

A DNA molecule of the present invention can be  
synthesized according to the phosphotriester method [Matteucci  
et al., JACS, 103: 3185 (1988)]. This method is particularly  
25 suitable when it is desired to effect site-directed  
mutagenesis of an instant DNA sequence, whereby a desired  
nucleotide substitution can be readily made. Another method  
for making an instant DNA molecule is by simply growing cells



transformed with plasmids containing the DNA sequence, lysing the cells, and isolating the plasmid DNA molecules.

Preferably, an isolated DNA molecule of the invention is made by employing the polymerase chain reaction (PCR) [e.g., U.S.

5 Patent No. 4,683,202 (issued to Mullis)] using synthetic primers that anneal to the desired DNA sequence, whereby DNA molecules containing the desired nucleotide sequence are amplified. Also, a combination of the above methods can be employed, such as one in which synthetic DNA is ligated to  
10 CDNA to produce a quasi-synthetic gene [e.g., U.S. Patent No. 4,601,980 (issued to Goeddel et al.)].

A further aspect of the invention is for a vector, e.g., a plasmid, that contains at least one of the nucleotide sequences shown in SEQ ID Nos. 1-4 or those predicted by the  
15 genomic clone in SEQ ID No. 21. Whenever the reading frame of the vector is appropriately selected, the vector encodes an IR polypeptide of the invention. Hence, as well as full-length protein, fragments of the native IR protein are contemplated; as well as fusion proteins that incorporate an amino acid  
20 sequence as described herein. Also, a vector containing a nucleotide sequence having a high degree of homology with any of SEQ ID Nos. 1-4 or 21 is contemplated within the invention, particularly when it encodes a protein having imidazoline binding activity.

25 A recombinant vector of the invention can be formed by ligating an afore-mentioned DNA molecule to a preselected expression plasmid, e.g., with T4 DNA ligase. Preferably, the plasmid and DNA molecule are provided with cohesive

(overlapping) termini, with the plasmid and DNA molecule operatively linked (i.e., in the correct reading frame).

Another aspect of the invention is a host cell transfected with a vector of the invention. Relatedly, a protein expressed by a host cell transfected with such a vector is contemplated, which protein may be bound to the cell membrane. Such a protein can be identical with an aforementioned polypeptide, or it can be a fragment thereof, such as when the polypeptide has been partially digested by a protease in the cell. Also, the expressed protein can differ from an aforementioned polypeptide, as whenever it has been subjected to one or more post-translational modifications. For the protein to be useful within the context of the present invention, it should exhibit imidazoline binding capacity.

A method of producing an imidazoline receptor protein is another aspect of the invention, which entails transfecting a host cell with an aforementioned vector, and culturing the transfected host cell in a culture medium to generate the imidazoline receptor. The receptor molecule can undergo any post-translational modification(s), including proteolytic decomposition, whereby its structure is altered from the basic amino acid residue sequence encoded by the vector. A suitable transfection method is electroporation, and the like.

With respect to transfecting a host cell with a vector of the invention, it is contemplated that a vector encoding an instant polypeptide can be transfected directly in animals. For instance, embryonic stem cells can be transfected, and the cells can be manipulated in embryos to produce transgenic

animals. Methods for performing such an operation have been previously described [Bond et al., Nature, 374:272-276 (1995)]. These methods for expressing an instant CDNA molecule in either tissue culture cells or in animals can be especially useful for drug discovery.

Possibly the most significant aspect of the present invention is in its potential for affording a method of screening for a ligand (drug) that binds to an imidazoline receptor. Such a method comprises culturing an above-mentioned host cell in a culture medium to express an instant imidazoline receptive polypeptide, then contacting the polypeptides with a labelled ligand, e.g., radiolabelled p-iodoclonidine, for the imidazoline receptor under conditions effective to bind the labelled ligand thereto. The polypeptides are further contacted with a candidate ligand, and any displacement of the labelled ligand from the polypeptides is detected. Displacement signifies that the candidate ligand actually binds to the imidazoline receptor. These steps could be performed on intact host cells, or on proteins isolated from the cell membranes of the host cells.

Typically, a suitable drug screening protocol involves preparing cells (or possibly tissues from transgenic animals) that express an instant imidazoline receptive polypeptide. In this process, categories of chemical structure are systematically screened for binding affinity or activation of the receptor molecule encoded by the transfected CDNA. This process is currently referred to as combinatorial chemistry. With respect to the imidazoline receptor, a number of

commercially available radioligands, e.g.,  $^{125}\text{PIC}$ , can be used for competitive drug binding affinity screening.

An alternative approach is to screen for drugs that elicit or block a second messenger effect known to be coupled to activation of the imidazoline receptor, e.g., moxonidine-stimulated arachidonic acid release. Even with a weak binding affinity or activation by one category of chemicals, systematic variations of that chemical structure can be studied and a preferred compound (drug) can be deduced as being a good pharmaceutical candidate. Identification of this compound would lead to animal testing and upwards to human trials. However, the initial rationale for drug discovery becomes vastly improved with an instant cloned imidazoline receptor.

Along these lines, a drug screening method is contemplated in which a host cell of the invention is cultured in a culture medium to express an instant imidazoline receptive polypeptide. Intact cells are then exposed to an identified agent (ie., agonist, inverse agonist, or antagonist) under conditions effective to elicit a second messenger or other detectable responses upon interacting with the receptor molecule. The imidazoline receptive polypeptides are then contacted with one or more candidate chemical compounds (drugs), and any modification in a second messenger response is detected. Compounds that mimic an identified agonist would be agonist candidates, and those producing the opposite response would be inverse agonist candidates. Those compounds that block the effects of a known agonist would be

antagonist candidates for an in vivo imidazoline receptor. For meaningful results, the contacting step with a candidate compound is preferably conducted at a plurality of candidate compound concentrations.

5           A method of probing for another gene encoding an imidazoline receptor or homologous protein is further contemplated. Such a method comprises providing a radiolabelled DNA molecule identical or complementary to one of the above-described CDNA molecules (probe). The probe is  
10 then placed in contact with genetic material suspected of containing a gene encoding an imidazoline receptor or encoding a homologous protein, under stringent hybridization conditions (e.g., a high stringency wash condition is 0.1 x SSC, 0.5% SDS at 65°C), and identifying any portion of the genetic material  
15 that hybridizes to the DNA molecule.

          Still further, a method of selectively producing antibodies, (e.g., monoclonal antibodies, immunoreactive with an instant imidazoline-receptive protein) comprises injecting a mammal with an aforementioned polypeptide, and isolating the  
20 antibodies produced by the mammal. This aspect is discussed in more detail in an example presented hereinafter.

          The present inventors began their search for a human imidazoline receptor CDNA by screening a  $\lambda$ gt11 phage human hippocampus CDNA expression library. Their research had  
25 indicated that both of the known antisera (Reis and Dontenwill) that are directed against human imidazoline receptors were immunoreactive with identical bands on SDS gels of membranes prepared from the human hippocampus (an in other

tissues). By contrast, other brain regions either were commercially unavailable as cDNA expression libraries or yielded inconsistencies between the two antisera. Therefore, it was felt that a human hippocampal cDNA library held the best opportunity for obtaining a cDNA for an imidazoline receptor. Immunoexpression screening was chosen over other cloning strategies because of its sensitivity when coupled with the ECL detection system used by the present inventors, as discussed hereinbelow.

A number of unique discoveries led to identifying the first 5A-1 clone as an imidazoline receptor cDNA. These included discoveries that led to the choice of a hippocampal cDNA library and adapting ECL to the antisera. Once the initial clone (5A-1) was identified and sequenced, a more complete clone (EST04033) was purchased without restriction from ATCC Inc. (Catalogue # 82815; American Type Culture Collection, Rockville, MD). EST 04033 was the only EST clone available at the time of the discovery of 5A-1, that contained a segment of complete homology (the origination of EST04033 is discussed later on). The binding affinities of the expressed protein after transfection in COS cells were determined by radioligand binding procedures developed in the inventor's laboratory [Piletz and Sletten, 1993, *ibid*].

To identify an instant cDNA clone encoding an imidazoline receptor it was preferable to employ both of the known antibodies to imidazoline receptors. These antibodies were obtained by contacting Dr. D. Reis (Cornell University Medical Center, New York City), and Drs. M. Dontenwill and P. Bousquet

(Laboratoire de Pharmacologie Cardiovascular et Renale, CNRS, Strasbourg, France). These antisera were obtained free of charge and without confidentiality or restrictions on their use. The former antiserum (Reis antiserum) was derived from a published imidazoline receptor protein [Wang et al., (1992, 1993), the disclosures of which are incorporated herein by reference]. The method for raising the latter antiserum (Dontenwill antiserum) has also been published [Bennai et al., (1995), the disclosure of which is also incorporated herein by reference]. The latter antiserum was developed using an anti-idiotypic approach that identified the pharmacologically correct (clonidine and idazoxan selective) binding site structure.

#### Example 1. Selectivity of the Antisera.

The obtained Reis antiserum had been prepared against a purified imidazoline binding protein isolated from BAC cells, which protein runs in denaturing-SDS gels at 70 Kda [Wang et al., 1992, 1993]. The Dontenwill antiserum is anti-idiotypic, and thus is believed to detect the molecular configuration of an imidazoline binding site domain in any species. Prior to being used for screening plaques, both antisera were cleaned by stripping out possible antibacterial antibodies.

Both antisera have been tested to ensure that they are in fact selective for a human imidazoline receptor. In particular, we found that both of these antisera detected identical bands in human platelets and hippocampus, and in brainstem RVLM (NRL) by Western blotting (see Fig. 1). In

these studies, in order to increase sensitivity over previously published detection methods, an ECL (Enhanced Chemiluminescence) system was employed. The linearity of response of the ECL system was demonstrated with a standard curve. ECL detection was demonstrated to be very quantifiable and about ten times more sensitive than other screening methods previously used with these antisera. Western blots with antiserum dilutions of 1:3000 revealed immunoreactivity with as little as 1 ng of protein from a human hippocampal homogenate by dot blot analysis.

For the studies depicted in Fig. 1, human hippocampal homogenate (30 $\mu$ g) and NRL membrane proteins (10 $\mu$ g) were electrophoresed through a 12.5% SDS-polyacrylamide gel, electrotransferred to nitrocellulose and sequentially incubated with (1) the Reis antibody (1:2000 dilution) and (2) the Dontenwill antibody (1:5000 dilution). Immunoreactive bands were visualized with an Enhanced Chemiluminescence (ECL) detection kit (Amersham) using anti-rabbit Ig-HRP conjugated antibody at a dilution of 1:3000 and the ECL detection reagents. Following detection with the antibody, blots were stripped and reprocessed omitting the primary antibody to check for complete removal of this antibody. In panels A and B, lane 1 shows the immunoreactive bands observed with the Reis antibody and lane 2 shows the bands detected with the Dontenwill antibody. Protein molecular weight standards are indicated to the left of each panel (in Kda).

Despite the diverse origins of the Reis and Dontenwill antisera, both of these antisera detected a similar 85 Kda



protein in human brain and other tissues. But, a 33 Kda band was found in human platelets. Although the 33 Kda band is of smaller size than that reported for other tissues [Wang et al., 1993; Escriba et al., 1994; Grenney et al., 1994], the fact that both antisera detected it, suggests that both the 85 Kda and 33 Kda bands may be imidazoline binding polypeptides. The 85 and 33 Kda bands were enriched in plasma membrane fractions, as is known to be the case for  $IR_1$  binding, but not  $I_2$  binding [Piletz and Sletten, 1993].

A significant positive correlation was established for the 85 Kda band detected by the Dontenwill antiserum with  $IR_1$  Bmax values across nine rat tissues ( $r^2 = 0.8736$ ). A similar positive correlation was established amongst platelet samples from 15 healthy platelet donors between radioligand  $IR_1$  Bmax values (but not  $I_2$  or  $\alpha_2AR$  Bmax values), and the 33 Kda band (presumed  $IR_1$  immunoreactivity) on Western blots. This correlation exhibited a slope factor close to unity (results not shown). These correlations strongly suggested that an  $IR_1$  binding protein might be revealed in an imidazoline receptor-antibody Western blotting assay. Furthermore, the Reis antiserum failed to detect authentic  $\alpha_2AR$ , MAO A, or MAO B bands on gels, i.e., it was not immunoreactive with MAO at MW = 61 Kda, or  $\alpha_2AR$  at MW = 64 Kda. Additionally, no immunoreactive bands were observed using preimmune antiserum. Thus, after extensively characterizing these antisera with human and rat materials, it was concluded that these antisera are indeed selective for human imidazoline receptor protein.

Example 2. Cloning of cDNA For An Imidazoline Receptor

A commercially available human hippocampal cDNA  $\lambda$ gt11 expression library (Clontech Inc., Palo Alto, CA) was screened for immunoreactivity sequentially using both the anti-  
5 idiotypic Dontenwill antiserum and the Reis antiserum. Standard techniques to induce protein and transference to a nitrocellulose overlay were employed. [See, for instance, Sambrook et al., 1989, "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press]. After washing  
10 and blocking with 5% milk, the Dontenwill antiserum was added to the overlay at 1:20,000 dilution in Tris-buffered saline, 0.05% Tween20, and 5% milk. The Reis antiserum was employed similarly, but at 1:15,000 dilution. These high dilutions of primary antiserum were chosen to avoid false positives. The  
15 secondary antibody was added, and positive plaques were identified using ECL. Representative results are shown in Fig. 2.

Positive plaques were pulled and rescreened until tertiary screenings yielded only positive plaques. Four  
20 separate positive plaques were identified from more than 300,000 primary plaques in our library. Recombinant  $\lambda$ gt11 DNA purified from each of the four plaques was subsequently subcloned into *E. coli* pBluescript vector (Stratagene, La Jolla, CA). Sequencing of these four cDNA inserts in  
25 pBluescript demonstrated that they were identical, suggesting that only one cDNA had actually been identified four times. Thus, the screening had been verified as being highly reproducible and the frequency of occurrence was as expected

for a single copy gene, i.e., one in 75,000 transcripts. As shown in Fig. 2, the protein produced by the first positive clone, designated 5A-1, tested positive with both the Reis antiserum and the Dontenwill antiserum. Clone 5A-1 has been deposited under the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, MD, USA, 20852, on August 28, 1997 and has been assigned deposit accession no. ATCC 209217. Tertiary-screened plaques of 5A-1 were all immuno-positive with either of the two known anti-imidazoline receptor antisera, but not with either preimmune antisera. These results suggested that clone 5A-1 encoded a fusion peptide similar to or identical with one of the predominant bands detected in human Western blots by both the Dontenwill and Reis antisera.

Sequencing of the first four clones was performed by contracting with ACGT Company (Chicago, IL) after subcloning them into pBluescript vector SK (Stratagene). Both manual and automatic sequencing strategies were employed which are outlined as follows:

#### Manual Sequencing

1. DNA sequencing was performed using T7 DNA polymerase and the dideoxy nucleotide termination reaction.

2. The primer walking method [Sambrook et al., *ibid.*] was used in both directions.

3. (<sup>35</sup>S)dATP was used for labelling.

4. The reactions were analyzed on 6% polyacrylamide wedge or non-wedge gels containing 8 M urea, with samples being loaded in the order of A C G T.

5. DNA sequences were analyzed by MacVector Version 5.0. and by various Internet-available programs, i.e., the BLAST program.

#### Automatic Sequencing

5 1. DNA sequencing was performed by the fluorescent dye terminator labelling method using AmpliTaq DNA polymerase (Applied Biosystems Inc., Prizm DNA Sequencing Kit, Perkin-Elmer Corp., Foster City, CA).

10 2. The primer walking method was used. The primers actually used were a subset of those shown in SEQ ID Nos. 7-20.

3. Sequencing reactions were analyzed on an Applied Biosystems, Inc. (Foster City, CA) sequence analyzer.

15 These results demonstrated that the initial clone (5A-1) contained a 1171 base pair insert (see SEQ ID No. 4). The entire 5A-1 cDNA was found to exist as extended open reading frame for translation into protein. Consequently, it was determined that the 5A-1 cDNA must be a fragment of a larger mRNA.

#### 20 cDNA Sequence Homologies

Using programs and databases available on the Internet (retrieved from NCBI Blast E-mail Server address blast@ncbi.nlm.nih.gov), it was determined that the 5A-1 clone encodes a previously undefined unique molecule. The BLASTP  
25 program [1.4.8MP, 20-June-1995 (build 11/13/95)] was used to compare all of the possible frames of amino acid sequences encoded by 5A-1 versus all known amino acid sequences

available within multiple international databases [Altschul et al., J. Mol. Biol., 215: 403-410 (1990)]. Only one protein, from *Micrococcus luteus*, possessed a marginally significant homology ( $p=0.04$ ) (41%) over a short stretch of 75 of the 390 amino acids encoded by 5A-1. Otherwise, there were not any amino acid homologies (i.e., with  $p \leq 0.05$ ) for any known proteins. Therefore, the protein encoded by 5A-1 is not significantly related to MAO A or B,  $\alpha_2AR$ , or any other known eukaryotic protein in the literature.

In contrast to the amino acid search on BLASTP, two nearly homologous EST cDNA sequences of undefined nature covering 155 and 250 b.p. of the 5A-1 clone were reported to exist using BLASTN (reached from the same Internet server on 11/13/95). BLASTN is a program used to compare known DNA sequences from international databases, regardless of whether they encode a polypeptide. Neither of the two EST cDNA sequences having high homology to 5A-1, to our knowledge have been reported anywhere else except on the Internet. Both were derived as Expressed Sequence Tags (ESTs) in random attempts to sequence the human cDNA repertoire [as described in Adams et al., Science, 252: 1651-1656 (1991)]. As far as can be determined, the people who generated these ESTs lack any knowledge of what protein(s) they encode. One cDNA, designated HSA09H122, contained 250 b.p. with 7 unknown/incorrect base pairs (97% homology) versus 5A-1 over the same region. HSA09H122 was generated in France (Genethon, B.P. 60, 91002 Evry Cedex France) from a human lymphoblast cDNA library. The other EST, designated EST04033, contained

155 b.p. with 12 unknown/incorrect base pairs (92% homology) versus 5A-1 over the same region. EST04033 was generated at the Institute for Genomic Research (Gaithersburg, MD) from a human fetal brain cDNA clone (HFBDP28). Thus, both of these  
5 ESTs are short DNA sequences and contain a number of errors (typical of single-stranded sequencing procedures as used when randomly screening ESTs).

Based on the BLASTN search, the owner of HSA09H122 was contacted in an effort to obtain that clone. The current  
10 owner of the clone appears to be Dr. Charles Auffret (Paul Brousse Hospital, Genetique, B.P. 8, 94801 Villejuif Cedex, France). Dr. Auffret indicated by telephone that his clone came from a lot of clones believed to be contaminated with yeast DNA, and he did not trust it for release. Contamination  
15 with yeast DNA of that clone was later confirmed to have been reported within an Internet database. Thus, HSA09H122 was not reliable.

The other partial clone (EST04033) was purchased from American Type Culture Collection in Rockville, MD (ATCC  
20 Catalog no. 82815). A telephone call to the Institute for Genomic Research revealed that it had been deposited at ATCC under [insert terms]. As far as can be determined, the present inventors were the first to completely sequence EST04033. The complete size of EST04033 was 3389 b.p. (SEQ ID No. 1), with a  
25 3,318 b.p. nonplasmid insert (see SEQ ID No. 3). Within this sequence of EST04033 the remaining 783 base pairs of the coding sequence presumed for a 70 kDa imidazoline receptor were predicted at the 5' side of 5A-1 (i.e., 783 coding

nucleotides unique to EST04033 + 1171 coding nucleotides of 5A-1 = 1954 predicted total coding nucleotides; assuming b.p.# 1397-1400 in SEQ. No. 1 encodes the initiating methionine). The entire 1954 b.p. coding region for an  $\approx$  70 kDa protein is shown in SEQ ID No. 2. The nucleotide sequence of EST04033 was determined in the same manner as described previously for the 5A-1 clone. The nucleotide sequence of the entire clone is shown in SEQ ID No. 1. In this sequence, an identical overlap was observed for the sequence obtained previously for the 5A-1 clone and the sequence obtained for EST04033. The 5A-1 overlap began at EST04033 b.p. 2,181 (SEQ. No.1) and continued to the end of the molecule (b.p. 3,351).

#### Conclusions About Our cDNA Clones

cDNA of the present invention encode a protein that is immunoreactive with both of the known selective antisera for an imidazoline receptor, i.e., Reis antiserum and Dontenwill antiserum. Thus, an instant cDNA molecule produces a protein immunologically related to a purified imidazoline receptor and has the antigenic specificity expected for an imidazoline binding site. These antisera have been documented in the scientific literature as being selective for an "imidazoline receptor", which provides strong evidence that such an imidazoline receptor has indeed been cloned.

As mentioned, our instant cDNA sequence contains open reading frame distinct from any previously described proteins. Therefore, the encoded protein is novel, and it is unrelated to  $\alpha_2$ -adrenoceptors or monoamine oxidases. Small hydrophobic

domains in the predicted amino acid sequence suggest that the protein is probably membrane bound, as expected for an imidazoline receptor.

### Example 3. Cloning of a Human Gene

5           A pre-made genomic library of human placental DNA was purchased from Stratagene (La Jolla, CA) to screen for an IR gene by hybridization. The genomic library was constructed in Stratagene's vector  $\lambda$  FIX® II (catalog # 946206), and it was grown in XL1-Blue MRA (P2) host bacteria. It was titered to  
10           yield approximately 50,000 plaques per 137 mm plate. Lifts from six such plates were screened in duplicate by hybridization. The DNA probe used for screening was a 1.85 kb EcoRI fragment from EST 04033 cDNA (uniquely related to our sequences based on the BLASTN). After the restriction  
15           digestion of EST 04033 DNA, the 1.85 kb fragment was extracted from an agarose electrophoresis gel, cleaned according to the GENECLAN® III kit manual (BIO 101, Inc., P.O. Box 2284, La Jolla, CA), and radiolabeled with [ $\alpha$ -<sup>32</sup>P]d-CTP according to Stratagene's Prime-It® II Random Primer Labeling Kit manual.  
20           Plaques were lifted onto 137 mm Duralon-UV™ membranes (Stratagene's catalog #420102), denatured, and cross-linked with Stratagene's UV-Stratalinker™ 1800. Hybridization was conducted under high stringency conditions: prehybridization = 6 X SSC, 1 % SDS, 50 % formamide, and 100  $\mu$ g/ml of sheared,  
25           denatured salmon sperm DNA at 42°C for 2 hrs; hybridization = 6 X SSC, 1 % SDS, 50 % formamide, and 100  $\mu$ g/ml of sheared, denatured salmon sperm DNA at 45°C overnight; wash = 2 washes



of 1 X SSC, 0.1 % SDS at 65°C and 3 washes of 0.2 X SSC, 0.2 % SDS at 65°C. This hybridization procedure is essentially described in Stratagene's vector  $\lambda$  FIX® II instruction manual. Positive plaques were localized by developing Kodak BioMax films. Two positive genomic clones of identical size were retained through three rounds of screening.

One of the positive genomic clones (designated JEP 1-A) was selected for complete characterization. It was found to contain an  $\approx$  17 kb insert. Large-scale preparations of this genomic clone DNA were performed using the  $\lambda$  QUICK! SPIN kit (BIO101, La Jolla, CA). To verify that we had cloned a gene corresponding to 5A-1 and EST04033 cDNA, some restriction site positions in the genomic clone were determined using the FLASH Nonradioactive Gene Mapping Kit (Stratagene) and compared to Southern blots of human DNA. The location of genomic sequences highly related to (or identical to) those of our cDNA clones was determined by high stringency hybridization (as above) with the following  $^{32}\text{P}$ -labeled probe: a 1110 bp *Apa*I-*Eco*RI fragment from the cDNA clone 5A-1. This fragment was chosen as the probe because it lacks the GAG repeat (encoding glutamic acids), which might have complicated matters if it were found to be repeated elsewhere in the genome. With genomic clone JEP1-A, we detected a 14.1 kb *Eco*RI fragment and a 7.7 kb *Sac*I fragment that hybridized with this probe. Southern blots containing *Eco*RI- or *Sac*I-digested human genomic DNA (from human blood) with the 1110 bp *Apa*I-*Eco*RI cDNA probe also resulted in the detection of a 14.1 kb *Eco*RI fragment and a 7.7 kb *Sac*I fragment. No additional

restriction fragments of human genomic DNA appeared to hybridize with this probe under lower stringency conditions. These results strongly suggested that this gene (JEP-1A) encodes transcript(s) giving rise to the 5A-1 and EST04033 cDNA clones. Clone JEP-1A has been deposited under the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, MD, USA, 20852, on August 28, 1997 and has been assigned deposit accession no. ATCC 209216.

Genomic DNA sequencing was done by contract with Cadus Pharmaceutical Corporation (Tarrytown, NY). The original lambda JEP1-A clone was subcloned into pZero (Invitrogen) as a convenient vector. The initial fragments for sequencing were derived from Sac I and Xba I restriction enzymes. The short Sac I fragments of 1.5, 3.0 and 3.5 kb were further digested with Hind III, Pst I, and Kpn I yielding 15 subclones of varying length. The procedure consisted of sequencing all these subclones and parent clones with vector forward and reverse primers. Subsequently, this initial round of sequencing was supplemented with primer walking using custom oligonucleotides. The Sac I fragments were joined together by primer walking using the 2 Xba I fragments of 3 and 10 Kb. Then, the largest Sac I fragment (8 kb) and the 10 kb Xba I fragment were used as templates for a transposon sequencing method. The method used was the Primer Island Transposition Kit (Perkin-Elmer Corp., Norwalk, CT; Applied Biosystems) (ABI). The kit consists of a synthetic transposon (Ty1) containing forward and reverse primers and the integrase

enzyme which inserts the transposon randomly into the target plasmid DNA. Transposon insertion is an alternative to subcloning or primer walking when sequencing a large region of DNA (Devine and Boeke, Nucleic Acids Res. 22: 3765-3772 (1994); Devine et al., Genome Res., in press, (1997); Kimmel et al., In Genome Analysis, a Laboratory Manual, Cold Spring Harbor Press, NY, NY, in press (1997). A total of over 250 individual sequencing reactions were performed. Sequencing was done on ABI model 373 and 377 automated sequencers using ABI dye-terminator sequencing kits. Primers were designed using Gene Runner software (Hastings Software, Hastings On Hudson, NY). Oligonucleotides were purchased from Gibco-BRL (Gaithersburg, MD). Sequence assembly was performed using Sequencer Software (Gene Codes Corp., Ann Arbor, MI) from 4-fold redundancy of sequences.

The entire sequence of our JEP-1A genomic clone is shown in SEQ. 21. The computer program, GENSCAN 1.0, was able to identify splice sites of known topology. As expected, this gene contained a number of introns. See Table 1 hereinbelow. Only one continuous open reading frame was identified within our genomic clone. This open reading frame was interrupted by a number of introns (which is typical of eukaryotic transcripts) as shown in Fig. 5. The predicted polypeptide is encoded by the genomic DNA beginning at b.p. # 971 of SEQ ID No. 21. The predicted amino acid sequence of the polypeptide encoded thereby is shown in SEQ ID No. 22. As can be seen, the entire 5A-1 DNA and polypeptide sequence was encapsulated within this predicted genomic transcript. Therefore, there is

no question that this is the gene encoding 5A-1 and EST04033 cDNA. In addition, JEP-1A has more nearly defined the full-length transcript (by at least 102 more coding nucleotides than the cDNAs alone).

TABLE 1

## Position of Predicted Introns and Exons

GENSCAN 1.0 Date run: 26-Aug-97 Time: 12:35:39  
Sequence gs\_seqfile : 15202 bp : 58.36% C+G : Isochore 4 (57.00 - 100.00 C+G%)

Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P..	Tscr...
1.01	Intr	+	971	1084	114	1	0	69	98	200	0.836	20.91
1.02	Intr	+	4096	4177	82	0	1	37	53	81	0.358	-0.13
1.03	Intr	+	5732	5856	125	0	2	117	95	84	0.953	13.48
1.04	Intr	+	6997	7046	50	0	2	95	116	44	0.998	6.52
1.05	Intr	+	8416	9825	1410	1	0	96	94	2914	0.970	283.09
1.06	Intr	+	10489	10897	409	1	1	15	59	318	0.517	17.19
1.07	Intr	+	11293	11449	157	0	1	57	61	236	0.998	18.57
1.08	Intr	+	11923	12051	129	2	0	84	63	224	0.997	21.34
1.09	Intr	+	12570	12731	162	1	0	95	80	229	0.996	23.94
1.10	Term	+	13090	13700	611	2	2	59	41	1012	0.942	89.44
1.11	PlyA	+	14257	14262	6							1.05

A BLASTN analysis of the entire genomic sequence (on 08/26/97) demonstrated again that this gene has not been previously defined in the literature.

As with the cDNA clones, some EST sequences of identity were found (listed below and later). Of particular interest was a variance in the first intron splice site predicted by the computer. Upstream of that site (ie., upstream of amino acids PEKKGGE = amino acids predicted after first splice site) we have identified two types of transcripts. Genomic clone JEP-1A predicted 34 amino acids upstream of that sequence before entering another intron upstream. In an identical

manner, three ESTs (H61282, AA428790 and AA428250) overlapped that entire region in our clones and they contained the identical nucleotides for those 34 amino acids, plus an additional 22 more amino acids further upstream. By

5 comparison, however, our EST04033 varied from all of these ESTs upstream of that site. This means, the first 1,532 nucleotides of EST04033 (thought to encode translation of amino acids 1-56 of EST04033 beginning at b.p. 1,398 in SEQ. 1) are completely at variance with the other ESTs down to that  
10 splice site, but from there on they are identical. This provides strong evidence that this site can generate two alternatively spliced transcripts which can produce at least one functional protein (ie., the transfections with EST04033). For the reader's information, this splice site is upstream of  
15 b.p. # 1,565 in SEQ.1, b.p. # 168 in SEQ.2, b.p. # 1,532 in SEQ.3, amino acid # 57 in SEQ.5, and b.p. # 971 in the genomic SEQ.21.

#### Genomic Sequence Analysis

Of interest is a unique glutamic- and aspartic acid-rich  
20 region within our predicted protein. This region of the IR protein delineates a highly unique span of 59 amino acids, 36 of which are Glu or Asp residues (61%). This region was largely discovered within clone 5A-1 and it is present within all discovered and predicted transcripts from the gene  
25 (EST04033 included). This sequence lies between two potential transmembrane loops (hydrophobic domains). The identification of this unique Glu/Asp-rich domain within our

clones is consistent with an expected negatively charged pocket capable of binding clonidine and agmatine, both of which are highly positively charged ligands. Also, since the Dontenwill antiserum was specifically developed against an idazoxan/clonidine binding site, and its immunoreactivity is directed against the clone 5A-1/ $\lambda$ gt11 fusion protein, this suggests that clone 5A-1 might encode an imidazoline binding site. Furthermore, this glu/asp-rich sequence is located within the longest stretch of homology that the clone has with any known protein, i.e., the ryanodine receptor (as determined by on BLASTN). Specifically, we have discovered four regions of homology between the imidazoline receptor and the ryanodine receptor, which are all Glu/Asp-rich. The total nucleic acid homology is 67% with the ryanodine receptor DNA over the stretches encompassing this region. However, this is not sufficient to indicate that the imidazoline receptor is a subtype of the ryanodine receptor, because this homologous stretch is still a minor portion of the overall transcript(s) identified in the gene. Instead, this significant homology may reflect a commonality in function between this region of the IR and the ryanodine receptor.

The Glu/Asp-rich region within the ryanodine receptor has also been reported to define a calcium and ruthenium red dye binding domain that modulates the ryanodine receptor/ $\text{Ca}^{++}$  release channel located within the sarcoplasmic reticulum. The only other charged amino acids within the Glu/Asp-rich region of our clones are two arginines (the ryanodine receptor has uncharged amino acids at the corresponding positions).

Based on this identification of Arg residues within the Glu/Asp-rich region of the predicted imidazoline binding site, the assistance of Dr. Paul Ernsberger (Case Western Reserve University, Cleveland, Ohio) was enlisted. Dr. Ernsberger performed phenylglyoxal attack of arginine on native PC-12 membranes. Dr. Ernsberger was able to demonstrate that this treatment completely eliminated imidazoline binding sites in these membranes. This provides some indirect evidence that the native imidazoline binding site also contains an Arg residue. On the other hand, attempts to chemically modify cysteine and tyrosine residues, which are not located near the Glu/Asp-rich region did not affect PC-12 membrane binding of imidazolines.

As a further test of the sequence, it was determined whether native IR binding sites in PC-12 cells would be sensitive to ruthenium red. From the structure of the cloned sequence, it was reasoned that native IR should bind ruthenium red. Accordingly, a competition of ruthenium red with  $^{125}\text{PIC}$  at native IR sites on PC-12 membranes was studied. In these studies it was observed that ruthenium red competed for  $^{125}\text{PIC}$  binding to the same extent as did the potent imidazoline compound, moxonidine, i.e., 100% competition. Furthermore, the  $\text{IC}_{50}$  for competition of ruthenium red at IR was slightly more robust than reported for ruthenium red on the activation of calcium-dependent cyclic nucleotide phosphodiesterase - the previous most potent effect of ruthenium red on any biological site - indicating possible pharmacological importance. It is also noteworthy that calcium failed to compete for  $^{125}\text{PIC}$

binding at PC-12 IR sites (as did a calcium substitute, lanthanum). We and others have previously reported that a number of other cations robustly interfere with IR binding [Ernsberger et al., Annals NY Acad.Sci., 763: 22-42 (1995);  
5 Ernsberger et al., Annals NY Acad.Sci., 763: 163-168 (1995)]. Attempts were also made to directly stain the proteins in SDS gels with ruthenium red [Chen and MacLennan, J. Biol. Chem., 269: 22698-22704 (1994)]. It was found that ruthenium red stains the same platelet (33 kDa) and brain (85 kDa) bands  
10 that Reis antiserum detects. (Remember, the same 33 kDa band was verified to directly correlate with <sup>125</sup>PIC Bmax values for IR.) Thus, these results linked the attributes predicted from the cloned sequence to a native IR binding site.

Two other facets of the predicted polypeptide from JEP-1A  
15 suggest that we have identified most of the functional sequences. First, our predicted protein is comparable in regard to both the order and size of three regions of importance to the function of the interleukin-2R $\beta$  receptor (IL-2R $\beta$ ). Specifically, IL-2R $\beta$  possesses the following  
20 regions over a span of 286 amino acids: ser-rich region, followed by glu/asp-rich region, followed by proline-rich region. Likewise, our predicted protein has the same three regions, in the same order, over a span of about 625 amino acids. This suggests that our protein might function  
25 similarly as cytokine receptors. Secondly, our predicted protein possesses a cytochrome p450 heme-iron ligand signature sequence [Nelson et al., Pharmacogenetics 6: 1-42 (1996)]. This suggests that our protein might also function as do



cytochrome p450s in oxidative, peroxidative and reductive metabolism of endogenous compounds.

Some additional findings about the amino acid sequence of our instant IR polypeptide are: (1) The glu/asp-rich region also bears similarity to an amino acid sequence within a GTPase activator protein. (2) There appear to be four small hydrophobic domains indicative of transmembrane domain receptors. (3) A number of potential protein kinase C (PKC) phosphorylation sites appear near to the carboxy side of the protein, and we have previously found that treatment of membranes with PKC leads to an enhancement of native IR binding. Thus, these observations are all consistent with other observations expected for native IR.

#### RNA Studies

Northern blotting has also been performed on polyA<sup>+</sup> mRNA from human tissues in order to ascertain the regional expression of the mRNA corresponding to our cDNA. The same 1110 b.p. *ApaI-EcoRI* fragment from cDNA clone 5A-1 used in Southern blots was used for these studies. This probe region was not found within any other known sequences on the BLASTN database. The results revealed a 6 kb mRNA band, which predominated over a much fainter 9.5 kb mRNA in most regions (Fig. 6). Some exceptions to this pattern were in lymph nodes and cerebellum (Fig. 6), where the 9.5 kb band was equally or more intense. Although the 6 kb band is weakly detectable in some non-CNS tissues, it is enriched in brain. An enrichment of the 6 kb mRNA is observed in brainstem, although not

exclusively. The regional distribution of the mRNA is somewhat in keeping with the reported distribution of IR binding sites, when extrapolated across species (Fig. 6). Thus, the rank order of Bmax values for IR in rat brain has been reported to be frontal cortex > hippocampus > medulla oblongata > cerebellum [Kamisaki et al., Brain Res., 514: 15-21 (1990)]. Therefore, with the exception of human cerebellum, which showed two mRNA bands, the distribution of the mRNA for our the present cloned cDNA is consistent with it belonging to IR.

[It should be noted that while IR binding sites are commonly considered to be low in cerebral cortex compared to brainstem, this is in fact a misinterpretation of the literature based only on comparisons to the alpha-2 adrenoceptor's Bmax, rather than on absolute values. Thus, IR Bmax values have actually been reported to be slightly higher in the cortex than the brainstem, but they only "appear" to be low in the cortex in comparison to the abundance of alpha-2 binding sites in cortex. Therefore, the distribution of the IR mRNA is reasonably in keeping with the actual Bmax values for radioligand binding to the receptor [Kamisaki et al., (1990)].

A final point to emphasize about the Northern blots is that they clearly demonstrate two high-stringency transcripts (Fig. 6). This is in keeping with the alternatively spliced EST cDNAs mentioned earlier. Thus, we suggest this may be the basis for both the 6 and 9.5 kb transcripts.

#### Summary of Genomic Sequence Results

The JEP-1A clone clearly contains most of the gene. Within it we have identified at least 3,776 nucleotides for transcript(s) (encoding 1,065 amino acids plus 587 b.p. of untranslated region down to the polyT<sup>+</sup> tail). This has been  
5 lengthened by at least 66 coding nucleotides upstream (22 amino acids) in comparison to overlapping ESTs. In addition to this, we are quite confident of the splice site for the two observed mRNA sizes. Most of the functional sequences are predicted to be encoded within our genomic clone.

10 A summary of the evidence that a gene encoding an imidazoline receptor protein has been cloned is summarized in Table 2 hereinbelow.

TABLE 2

Comparison of Protein Predicted From Our Clones with  
Properties of Native IR<sub>1</sub> and I<sub>2</sub> Sites

Imidazoline Receptor-like Clone	Authentic IR <sub>1</sub>	Authentic I <sub>2</sub>
Original $\lambda$ phage fusion protein (from 5A-1) is immunoreactive with Dontenwill and Reis antibodies	Dontenwill-Ab activity (a) inhibits RVLM IR <sub>1</sub> binding ( <sup>3</sup> H-Clonidine), & (b) correlates with 85 kDa Western band. Reis-Ab activity correlates w platelet IR <sub>1</sub> Bmax ( <sup>125</sup> PIC binding)	Dontenwill & Reis Abs both inhibit brain I <sub>2</sub> sites ( <sup>3</sup> H-IDX).
Segment homologous to a GTPase-activator prot'n	Weak to moderate sensitivity to GTP	Not sensitive to GTP
Predicts $\geq$ 120,000 MW protein	85,000 MW immunoreactivity	59-61,000 MW photoaffinity
Predicts 1-4 hydrophobic domains	Enriched in plasma membranes	Enriched in mitochondria
Encodes Glu/Asp-rich (negatively charged) domain consistent with Ca <sup>++</sup> and ruthenium red binding	<ul style="list-style-type: none"> <li>• Binds (+)-charged imidazolines</li> <li>• Sensitive to divalent cations</li> <li>• Sensitive to ruthenium red</li> </ul>	<ul style="list-style-type: none"> <li>• Binds (+)-charged imidazolines</li> <li>• Not sensitive to divalent cations</li> <li>• Unknown sensitivity for Ruthen. red</li> </ul>
Arginine is only positively charged amino acid near Glu/Asp domain	<ul style="list-style-type: none"> <li>• Arg attack eliminates binding</li> <li>• Cys &amp; Tyr attack w/o effect on binding</li> </ul>	Unknown
Encodes PKC sites	PKC treatment enhances binding	Unknown
Human mRNA Distribution; F.Cortex > hippocampus > medulla	Rat IR <sub>1</sub> Bmax ( <sup>125</sup> PIC): F.Cortex > hippocampus > medulla	Rat I <sub>2</sub> Bmax ( <sup>3</sup> H-IDX): Medulla > F. Cortex
Transfected COS-7 cells expressed high affinity for moxonidine & p-iodoclonidine (PIC)	High affinity for moxonidine and PIC	Low affinity for moxonidine and PIC

Example 4. Transient Transfection Studies

COS-7 cells were transfected with a vector containing EST04033 cDNA, which was predicted based on sequence analysis to contain the glu/asp rich region thought to be important for ligand binding to the imidazoline receptor protein. The EST04033 cDNA was subcloned into pSVK3 (Pharmacia LKB Biotechnology, Piscataway, NJ) using standard techniques [Sambrook, supra], and transfected via the DEAE-dextran technique as previously described [Choudhary et al., Mol.Pharmacol., 42: 627-633 (1992); Choudhary et al., Mol.Pharmacol., 43: 557-561 (1993); Kohen et al., J.Neurochem., 66: 47-56 (1996)]. A restriction map of the EST04033 cDNA is shown in Fig. 3. The restriction enzymes Sal I and Xba I were used for subcloning into pSVK3.

Briefly stated, COS-7 cells were seeded at  $3 \times 10^6$  cells/100 mm plate, grown overnight and exposed to 2 ml of DEAE-dextran/plasmid mixture. After a 10-15 min. exposure, 20 ml of complete medium (10% fetal calf serum; 5  $\mu$ g/ml streptomycin; 100 units/ml penicillin, high glucose, Dulbeccos' modified Eagle's medium) containing 80  $\mu$ M chloroquine was added and the incubation continued for 2.5 hr. at 37°C in a 5% CO<sub>2</sub> incubator. The mixture was then aspirated and 10 ml of complete medium containing 10% dimethyl sulfoxide was added with shaking for 150 seconds.

Following aspiration, 15 ml of complete medium with dialyzed serum was added and the incubation continued for an additional 65 hours. After this time period, the cells from 6 plates were harvested and membranes were prepared as

previously described [Ernsberger et al., Annals NY Acad. Sci., 763: 22-42 (1995), the disclosure of which is incorporated herein by reference]. Parent, untransfected COS-7 cells were prepared as a negative control. Some membranes were treated  
5 with and without PKC for 2 hrs prior to analysis, since previous studies had indicated that receptor phosphorylation could be beneficial to detect IR binding.

Transfected samples were also analyzed by Western blots. The protocol used for Western blot assay of transfected cells  
10 is as follows. Cell membranes were prepared in a special cocktail of protease inhibitors (1 mM EDTA, 0.1 mM EGTA, 1 mM phenylmethyl-sulfonylfluoride, 10 mM  $\epsilon$ -aminocaproic acid, 0.1 mM benzamide, 0.1 mM benzamide-HCl, 0.1 mM phenanthroline, 10  $\mu$ g/ml pepstatin A, 5 mM iodoacetamide, 10  $\mu$ g/ml antipain, 10  
15  $\mu$ g/ml trypsin-chymotrypsin inhibitor, 10  $\mu$ g/ml leupeptin, and 1.67  $\mu$ g/ml calpain inhibitor) in 0.25 M sucrose, 1 mM MgCl<sub>2</sub>, 5 mM Tris, pH 7.4. Fifteen  $\mu$ g of total protein were denatured and separated by SDS gel electrophoresis. Gels were equilibrated and electrotransferred to nitrocellulose  
20 membranes. Blots were then blocked with 10% milk in Tris-buffered saline with 0.1% Tween-20 (TBST) during 60 min. of gentle rocking. Afterwards, blots were incubated in anti-imidazoline receptor antiserum (1:3000 dil.) for 2 hours. Following the primary antibody, blots were washed and  
25 incubated with horseradish peroxidase-conjugated anti-rabbit goat IgG (1:3000 dil.) for 1 hr. Blots were extensively washed and incubated for 1 min. in a 1:1 mix of Amersham ECL detection solution. The blots were wrapped in cling-film

(SARAN WRAP) and exposed to Hyperfilm-ECL (Amersham) for 2 minutes. Quantitation was based on densitometry using a standard curve of known amounts of protein containing BAC membranes or platelet membranes run in each gel.

5        One nM [ $^{125}$ I]p-iodoclonidine was employed in the radioligand binding competition assays, since at this low concentration this radioligand is selective for the IR site much more than for  $I_2$  binding sites. The critical processes of membrane preparation of tissue culture cells and the  
10       radioligand binding assays of IR and  $I_2$  have been reviewed by Piletz and colleagues [Ernsberger et al., Annals NY Acad Sci., 763: 510-519 (1995)]. Total binding (n = 12 per experiment) was determined in the absence of added competitive ligands and nonspecific binding was determined in the presence of  $10^{-4}$  M  
15       moxonidine (n = 6 per experiment). Log normal competition curves were generated against unlabeled moxonidine, p-iodoclonidine, and (-) epinephrine. Each concentration of the competitors was determined in triplicate and the experiment was repeated thrice.

20       The protocol to fully characterize radioligand binding in the transfected cells entails the following. First, the presence of IR and/or  $I_2$  binding sites are scanned over a range of protein concentrations using a single concentration of [ $^{125}$ I]-p-iodoclonidine (1.0nM) and  $^3$ H-idazoxan (8nM),  
25       respectively. Then, rate of association binding experiments (under a  $10 \mu$ M mask of NE to remove  $\alpha_2$ AR interference) are performed to determine if the kinetic parameters are similar to those reported for native imidazoline receptors [Ernsberger

et al. Annals NY Acad. Sci., 763: 163-168 (1995)]. Then, full Scatchard plots of [ $^{125}$ I]-p-iodoclonidine (2-20 nM if like IR) and  $^3$ H-idazoxan (5-60 nM if like I<sub>2</sub>) binding are conducted under a 10  $\mu$ M mask of NE.

Total NE (10  $\mu$ M)-displaceable binding is ascertained as a control to rule out  $\alpha_2$ -adrenergic binding. The B<sub>max</sub> and K<sub>D</sub> parameters for the transfected cells are ascertained by computer modeling using the LIGAND program [McPherson, G., J.Pharmacol.Meth., 14: 213-228 (1985)] using 20  $\mu$ M moxonidine to define IR nonspecific binding, or 20  $\mu$ M cirazoline to define I<sub>2</sub> nonspecific binding.

The results of the transient transfection experiments of the imidazoline receptor vector into COS-7 cells are shown in Fig. 4. Competition binding experiments were performed using membrane preparations from these cells and  $^{125}$ PIC was used to radiolabel IR sites. A mask of 10  $\mu$ M norepinephrine was used to rule out any possible  $\alpha_2$ AR binding in each assay even though parent COS-7 cells lacked any  $\alpha_2$ AR sites. Moxonidine and p-iodoclonidine (PIC) were the compounds tested for their affinity to the membranes of transfected cells. As can be seen, the affinities of these compounds in competition with  $^{125}$ PIC were well within the high affinity (nM) range.

The following IC<sub>50</sub> values and Hill slopes were obtained in this study: moxonidine, IC<sub>50</sub> = 45.1 nM (Hill slope = 0.35  $\pm$  0.04); p-iodoclonidine without PKC pretreatment of the membranes, IC<sub>50</sub> = 2.3 nM (Hill slope = 0.42  $\pm$  0.06); p-iodoclonidine with PKC pretreatment of the membranes, IC<sub>50</sub> = 19.0 nM (Hill slope = 0.48  $\pm$  0.08). Shallow Hill slopes for [ $^{125}$ I]p-iodoclonidine have been reported before in studies of the interaction of moxonidine and p-iodoclonidine with the human platelet IR<sub>1</sub> binding site [Piletz



and Sletten, (1993)]. Epinephrine failed to displace any of the  
[<sup>125</sup>I]p-iodoclonidine binding in the transfected cells, as  
expected since this is a nonadrenergic imidazoline receptor.  
Furthermore, in untransfected cells less than 5% of the amount of  
5 displaceable binding was observed as for the transfected cells -  
and this "noise" in the parent cells all appeared to be low  
affinity (data not shown). These results thus demonstrate the  
high affinities of two imidazoline compounds, p-iodoclonidine and  
moxonidine, for the portion of our cloned receptor encoded within  
10 EST04033. PKC pretreatment of the membranes had no effect in the  
transfected COS cells.

It was also observed that the level of the expressed  
protein, as measured by Western blotting of the transfected  
cells, was consistent with the level of IR binding that was  
15 detected. In other words, a protein band was uniquely detected  
in the transfected cells, and it was of a density consistent with  
the amount of radioligand binding. Hence, the present results  
are in keeping with those expected for an imidazoline receptor.  
In summary, these data provide direct evidence that the EST04033  
20 clone encodes an imidazoline binding site having high affinities  
for moxonidine and p-iodoclonidine, which is expected for an IR  
protein.

#### Example 5. Stable Transfection Methods.

Stable transfections can be obtained by subcloning the  
25 imidazoline receptor cDNA into a suitable expression vector,

e.g., pRc/CMV (Invitrogen, San Diego, CA), which can then be used to transform host cells, e.g. CHO and HEK-293 cells, using the Lipofectin reagent (Gibco/BRL, Gaithersburg, MD) according to the manufacturer's instructions. These two host cell lines can be used to increase the permanence of expression of an instant clone. The inventors have previously ascertained that parent CHO cells lack both  $\alpha_2$ -adrenoceptor and IR binding sites [Piletz et al., J. Pharm. & Exper. Ther., 272: 581-587 (1995)], making them useful for these studies. Twenty-four hours after transfection, cells are split into culture dishes and grown in the presence of 600  $\mu$ g/ml G418-supplemented complete medium (Gibco/BRL). The medium is changed every 3 days and clones surviving in G418 are isolated and expanded for further investigation.

Example 6. Direct Cloning of More Complete Gene and Other Homologous Human IR.

Direct probing of other human genomic and cDNA libraries can be performed by preparing labelled cDNA probes from different subcloned regions of our clone. Commercially available human DNA libraries can be used. Besides the cDNA and genomic libraries we have already screened, another genomic library is EMBL (Clontech), which integrates genomic fragments up to 22 kbp long. It is reasonable to expect that introns may exist within other human IR genes so that only by obtaining overlapping clones can the full-length genes be sequenced. A probe encompassing the 5'

end of an instant cDNA is generally useful to obtain the gene promoter region. Clontech's Human PromoterFinder DNA Walking procedure provides a method for "walking" upstream or downstream from cloned sequences such as cDNAs into adjacent genomic DNA.

5     Example 7. Methods for Preparing Antibodies to Imidazoline Receptive Proteins.

          An instant imidazoline receptive polypeptide can also be used to prepare antibodies immunoreactive therewith. Thus, synthetic peptides (based on deduced amino acid sequences from  
10    the DNA) can be generated and used as immunogens. Additionally, transfected cell lines or other manipulations of the DNA sequence of an instant imidazoline receptor can provide a source of purified imidazoline receptor peptides in sufficient quantities for immunization, which can lead to a source of selective  
15    antibodies having potential commercial value.

          In addition, various kits for assaying imidazoline receptors can be developed that include either such antibodies or the purified imidazoline receptor protein. A purification protocol has already been published for the bovine imidazoline receptor in  
20    BAC cells [Wang et al, 1992] and an immunization protocol has also been published [Wang et al., 1993]. These same protocols can be utilized with little if any modification to afford purified human IR protein from transfected cells and to yield selective antibodies thereto.

25           In order to obtain antibodies to a subject peptide, the

peptide may be linked to a suitable soluble carrier to which antibodies are unlikely to be encountered in human serum.

Illustrative carriers include bovine serum albumin, keyhole limpet hemocyanin, and the like. The conjugated peptide is injected into a mouse, or other suitable animal, where an immune response is elicited. Monoclonal antibodies can be obtained from hybridomas formed by fusing spleen cells harvested from the animal and myeloma cells [see, e.g., Kohler and Milstein, Nature, 256: 495-497 (1975)].

Once an antibody is prepared (either polyclonal or monoclonal), procedures are well established in the literature, using other proteins, to develop either RIA or ELISA assays [see, e.g., "Radioimmunoassay of Gut Regulatory Peptides; Methods in Laboratory Medicine," Vol. 2, chapters 1 and 2, Praeger Scientific Press, 1982]. In the case of RIA, the purified protein can also be radiolabelled and used as a radioactive antigen tracer.

Currently available methods to assay imidazoline receptors are unsuitable for routine clinical use, and therefore the development of an assay kit in this manner could have significant market appeal. Suitable assay techniques can employ polyclonal or monoclonal antibodies, as has been previously described [U.S. Patent No. 4,376,110 (issued to David et al.), the disclosure of which is incorporated herein by reference].

Summary

In summary, we have identified unique DNA sequences that have properties expected of a gene and the cDNA transcript(s) of an imidazoline receptor. Prior to our first cloning the cDNA, only  
5 two sequences of EST cDNA were identified within public databases having similar nature. But, these were both partial and imprecise sequences - not identified at all with respect to any encoded protein. Indeed, one of them (HSA09H122) was reported to be contaminated. In our hands, the other EST 04033 clone was  
10 correctly sequenced for the first time (in its entirety = 3318 bp). Prior to this, even the size of EST 04033 was unknown. The present inventors also demonstrated that an imidazoline receptive site can be expressed in cells transfected with the EST 04033 cDNA clone, and this site has the proper potencies of an IR. We  
15 have deduced most of the complete cDNA encoding this protein.

The present invention has been described with reference to specific examples for purposes of clarity and explanation. Certain obvious modifications of the invention readily apparent to one skilled in the art can be practiced within the scope of  
20 the appended claims.

SEQUENCE LISTING

## GENERAL INFORMATION:

APPLICANT: John E. Piletz

Tina R. Ivanov

TITLE OF INVENTION: DNA MOLECULES ENCODING IMIDAZOLINE RECEPTIVE  
POLYPEPTIDES AND POLYPEPTIDES ENCODED THEREBY

NUMBER OF SEQUENCES: 22

## CORRESPONDENCE ADDRESS:

ADDRESSEE: WENDEROTH, LIND &amp; PONACK

STREET: 805 Fifteenth St. N.W., Suite 700

CITY: Washington

STATE: District of Columbia

COUNTRY: U.S.A.

ZIP: 20005

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.50" 1.44 Mb diskette

COMPUTER: IBM PC compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1+

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/15695

FILING DATE: September 3, 1997

CLASSIFICATION:

## PRIOR APPLICATION DATA:

APPLICATION NO.: USSN 60/012,600

FILING DATE: March 1, 1996

## PRIOR APPLICATION DATA:

APPLICATION NO.: USSN 08/650,766

FILING DATE: May 20, 1996

## ATTORNEY/AGENT INFORMATION:

NAME: Warren Cheek

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER: WMC-1342/clone

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-8850

TELEFAX: (202) 371-8856

INFORMATION FOR SEQ ID NO: 1

## SEQUENCE CHARACTERISTICS:

LENGTH: 3389 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

## ORIGINAL SOURCE:

ORGANISM: Homo sapiens

## IMMEDIATE SOURCE:

LIBRARY: cDNA

CLONE: EST04033 (HFBDP28)

## FEATURE:

NAME/KEY: predicted translation product when  
transfected

LOCATION: 1398 ... 3389

## SEQUENCE DESCRIPTION: SEQ ID NO: 1

GCTCTAGAAC TAGTGGATCC CCCGGGCTGC AGGAATTCCA GTTTAATACT AACCCTAATG	60
TGTGACTGCG GTTTACAAAG AGCTCTGTAT CACCTGGGAT AGCTTTCAGT AGCAATTCAC	120
TACAACTGGT CCTAAAAAAT AATAACAATA ATAATAATAA TTAGAGAATT AAAACCCAAC	180
AGCATGTTGA ATGGTTAAAA TCACGTAAGA ACTGAAATTT GGGGTGGGGG TGTCCCTCAAC	240
AGCTGAGCTT GTCCTAGCAG TGAAAATGCT CGCCTCCAAG CAGGGCTCAG AAAGGTCTGG	300
AGCCCTCCAG GCAGAGGGCT GAGCTCAGGG GGCTCTTGGA GGACACTCAC CCCATGGTCC	360
ATGGGATGCT TCTGGCTTCC TTAAAAACAG TTGGGCATCC GCATTGTATA AGTAGGTGGA	420
GACCCTAGTG TGTTCTTTT GAAGGATATG GGAAGGGAGG ATGACGAACT AGAGAAGTGG	480
GAGGGGACCA AAATCACTGA GGTCCCAGAA TATCATAGAT TTGGGTATAG GATTGGGGTC	540
ACTAAGAATT GAGCACCAGG AATTCCAGCT TCTTCCCATT AAAGAACTG GGACTGGTTT	600
TGCCTTGGAG GCCTATGTAG TGTTTTCTGC CCCTGTCCCA TACCAAGTCT CATTGATATT	660
TCTGCAGAA ATCAGATGAA AATCTATTTT TAAAGACCAT TGGGAGAATG GGTGGTGGAG	720
AAGGAGTTGG AGTGGGGTTG GGGGGCAGTT AAAAATGAAT AAAAATCTCT CAGCTACAGA	780
ACCCAAACAT CACTTCCCTC CGCATTCA CA GATTTCCCA GCAGTCCCCA GATGGTTGTT	840

TCCGTGGGGA CACAGCAGCT GCCTCATTTC CCTTCAGGCC CCATGGGCTG CTGGTCAACC 900  
 TCAGGATCTA CTAAAGATGA CGCAAATGCC GACTGAACAA TCTGAAACCC AAAGGACTCG 960  
 AGGAGAGACA TGTTCTGCTG AGGAGAGAAA GGTGAGCCAA GGGCAGGGCC CAGGTCCCCC 1020  
 AGGGGGCCCC CGAGAGCCCC GACATGCACC TTCTGGATGT GTTTGTTCAA GTAGGACTTA 1080  
 GAGCGGAAGA AGCTCCCACA TTCAGGGCAT GGGTACTTCT TCTCCCCATC AGACTCCATT 1140  
 TTGTTTTTGG GGA CTGCCAT GTCGCAGGAG AAAGAGCCAT TGGCACTCTG CTTCTCTGGC 1200  
 GTCTTCAGGT CGCTGGCATC TGAGAGGTCA CCATAGGAGT CAGAGCTCTC AATCGGATCC 1260  
 TGATGTGAGC ATTTCTGGCC TTCTCGGTTA CAGATACTGC AGAAGTTGCT GGGCCCCCTCG 1320  
 CTGTGCTTCT TCAGGTGGTC TGCCATGTAT GCTGCCCGCA AGTACTTCCC ACACACCTGG 1380  
 CAGGGCACCT TGTCTTC ATG ACA GGC CAG GTG GGA GCG CAG ACG GTC TCG 1430

Met Thr Gly Gln Val Gly Ala Gln Thr Val Ser

1 5 10

GGT GGC AAA AGA AGC ATT GCA GGT CTG ACA CTT GTG AGG CCG CTC AGA 1478  
 Gly Gly Lys Arg Ser Ile Ala Gly Leu Thr Leu Val Arg Pro Leu Arg

15 20 25

AGT GTG CAC CTG CTT GAT ATG TCC GTT CAA GTG ATC AGG CCT GGA GAA 1526  
 Ser Val His Leu Leu Asp Met Ser Val Gln Val Ile Arg Pro Gly Glu

30 35 40

GCC TTT CCC ACA GCT CTG GCA GAT GTA AGG CGG AAT TCC CCA GAG AAG 1574  
 Ala Phe Pro Thr Ala Leu Ala Asp Val Arg Trp Asn Ser Pro Glu Lys

45 50 55

AAG GGT GGT GAA GAC TCC CGG CTC TCA GCT GCC CCC TGC ATC AGA CCC 1622  
 Lys Gly Gly Glu Asp Ser Trp Leu Ser Ala Ala Pro Cys Ile Arg Pro

60 65 70 75

AGC AGC TCC CCT CCC ACT GTG GCT CCC GCA TCT GCC TCC CTG CCC CAG 1670  
 Ser Ser Ser Pro Pro Thr Val Ala Pro Ala Ser Ala Ser Leu Pro Gln

80 85 90

CCC ATC CTC TCT AAC CAA GGA ATC ATG TTC GTT CAG GAG GAG GCC CTG 1718



Pro Ile Leu Ser Asn Gln Gly Ile Met Phe Val Gln Glu Glu Ala Leu

95

100

105

GCC AGC AGC CTC TCG TCC ACT GAC AGT CTG ACT CCC GAG CAC CAG CCC 1766

Ala Ser Ser Leu Ser Ser Thr Asp Ser Leu Thr Pro Glu His Gln Pro

110

115

120

ATT GCC CAG GGA TGT TCT GAT TCC TTG GAG TCC ATC CCT GCG GGA CAG 1814

Ile Ala Gln Gly Cys Ser Asp Ser Leu Glu Ser Ile Pro Ala Gly Gln

125

130

135

GCA GCT TCC GAT GAT TTA AGG GAC GTG CCA GGA GCT GTT GGT GGT GCA 1862

Ala Ala Ser Asp Asp Leu Arg Asp Val Pro Gly Ala Val Gly Gly Ala

140

145

150

155

AGC CCA GAA CAT GCC GAG CCG GAG GTC CAG GTG GTG CCG GGG TCT GGC 1910

Ser Pro Glu His Ala Glu Pro Glu Val Gln Val Val Pro Gly Ser Gly

160

165

170

CAG ATC ATC TTC CTG CCC TTC ACC TGC ATT GGC TAC ACG GCC ACC AAT 1958

Gln Ile Ile Phe Leu Pro Phe Thr Cys Ile Gly Tyr Thr Ala Thr Asn

175

180

185

CAG GAC TTC ATC CAG CGC CTG AGC ACA CTG ATC CGG CAG GCC ATC GAG 2006

Gln Asp Phe Ile Gln Arg Leu Ser Thr Leu Ile Trp Gln Ala Ile Glu

190

195

200

CGG CAG CTG CCT GCC TGG ATC GAG GCT GCC AAC CAG CGG GAG GAG GGC 2054

Trp Gln Leu Pro Ala Trp Ile Glu Ala Ala Asn Gln Trp Glu Glu Gly

205

210

215

CAG GGT GAA CAG GGC GAG GAG GAG GAT GAG GAG GAG GAA GAA GAG GAG 2102

Gln Gly Glu Gln Gly Glu Glu Glu Asp Glu Glu Glu Glu Glu Glu Glu

220

225

230

235

GAC GTG GCT GAG AAC CGC TAC TTT GAA ATG GGG CCC CCA GAC GTG GAG 2150  
 Asp Val Ala Glu Asn Arg Tyr Phe Glu Met Gly Pro Pro Asp Val Glu  
 240 245 250

GAG GAG GAG GGA GGA GGC CAG GGG GAG GAA GAG GAG GAG GAA GAG GAG 2198  
 Glu Glu Glu Gly Gly Gly Gln Gly Glu Glu Glu Glu Glu Glu Glu Glu  
 255 260 265

GAT GAA GAG GCC GAG GAG GAG CGC CTG GCT CTG GAA TGG GCC CTG GGC 2246  
 Asp Glu Glu Ala Glu Glu Glu Arg Leu Ala Leu Glu Trp Ala Leu Gly  
 270 275 280

GCG GAC GAG GAC TTC CTG CTG GAG CAC ATC CGC ATC CTC AAG GTG CTG 2294  
 Ala Asp Glu Asp Phe Leu Leu Glu His Ile Arg Ile Leu Lys Val Leu  
 285 290 295

TGG TGC TTC CTG ATC CAT GTG CAG GGC AGT ATC CGC CAG TTC GCC GCC 2342  
 Trp Cys Phe Leu Ile His Val Gln Gly Ser Ile Arg Gln Phe Ala Ala  
 300 305 310 315

TGC CTT GTG CTC ACC GAC TTC GGC ATC GCA GTC TTC GAG ATC CCG CAC 2390  
 Cys Leu Val Leu Thr Asp Phe Gly Ile Ala Val Phe Glu Ile Pro His  
 320 325 330

CAG GAG TCT CGG GGC AGC AGC CAG CAC ATC CTC TCC TCC CTG CGC TTT 2438  
 Gln Glu Ser Trp Gly Ser Ser Gln His Ile Leu Ser Ser Leu Arg Phe  
 335 340 345

GTC TTT TGC TTC CCG CAT GGC GAC CTC ACC GAG TTT GGC TTC CTC ATG 2486  
 Val Phe Cys Phe Pro His Gly Asp Leu Thr Glu Phe Gly Phe Leu Met  
 350 355 360

CCG GAG CTG TGT CTG GTG CTC AAG GTA CGG CAC AGT GAG AAC ACG CTC 2534  
 Pro Glu Leu Cys Leu Val Leu Lys Val Arg His Ser Glu Asn Thr Leu

365	370	375	
TTC ATT ATC TCG GAC GCC GCC AAC CTG CAC GAG TTC CAC GCG GAC CTG			2582
Phe Ile Ile Ser Asp Ala Ala Asn Leu His Glu Phe His Ala Asp Leu			
380	385	390	395
CGC TCA TGC TTT GCA CCC CAG CAC ATG GCC ATG CTG TGT AGC CCC ATC			2630
Arg Ser Cys Phe Ala Pro Gln His Met Ala Met Leu Cys Ser Pro Ile			
400	405	410	
CTC TAC GGC AGC CAC ACC AGC CTG CAG GAG TTC CTG CGC CAG CTG CTC			2678
Leu Tyr Gly Ser His Thr Ser Leu Gln Glu Phe Leu Arg Gln Leu Leu			
415	420	425	
ACC TTC TAC AAG GTG GCT GGC GGC TGC CAG GAG CGC AGC CAG GGC TGC			2726
Thr Phe Tyr Lys Val Ala Gly Gly Cys Gln Glu Arg Ser Gln Gly Cys			
430	435	440	
TTC CCC GTC TAC CTG GTC TAC AGT GAC AAG CGC ATG GTG CAG ACG GCC			2774
Phe Pro Val Tyr Leu Val Tyr Ser Asp Lys Arg Met Val Gln Thr Ala			
445	450	455	
GCC GGG GAC TAC TCA GGC AAC ATC GAG TGG GCC AGC TGC ACA CTC TGT			2822
Ala Gly Asp Tyr Ser Gly Asn Ile Glu Trp Ala Ser Cys Thr Leu Cys			
460	465	470	475
TCA GCC GTG CGG CGC TCC TGC TGC GCG CCC TCT GAG GCC GTC AAG TCC			2870
Ser Ala Val Arg Arg Ser Cys Cys Ala Pro Ser Glu Ala Val Lys Ser			
480	485	490	
GCC GCC ATC CCC TAC TGG CTG TTG CTC ACG CCC CAG CAC CTC AAC GTC			2918
Ala Ala Ile Pro Tyr Trp Leu Leu Leu Thr Pro Gln His Leu Asn Val			
495	500	505	

ATC AAG GCC GAC TTC AAC CCC ATG CCC AAC CGT GGC ACC CAC AAC TGT 2966  
 Ile Lys Ala Asp Phe Asn Pro Met Pro Asn Arg Gly Thr His Asn Cys  
 510 515 520

CGC AAC CGC AAC AGC TTC AAG CTC AGC CGT GTG CCG CTC TCC ACC GTG 3014  
 Arg Asn Arg Asn Ser PHe Lys Leu Ser Arg Val Pro Leu Ser Thr Val  
 525 530 535

CTG CTG GAC CCC ACA CGC AGC TGT ACC CAG CCT CGG GGC GCC TTT GCT 3062  
 Leu Leu Asp Pro Thr Arg Ser Cys Thr Gln Pro Arg Gly Ala Phe Ala  
 540 545 550 555

GAT GGC CAC GTG CTA GAG CTG CTC GTG GGG TAC CGC TTT GTC ACT GCC 3110  
 Asp Gly His Val Leu Glu Leu Leu Val Gly Tyr Arg Phe Val Thr Ala  
 560 565 570

ATC TTC GTG CTG CCC CAC GAG AAG TTC CAC TTC CTG CGC GTC TAC AAC 3158  
 Ile Phe Val Leu Pro His Glu Lys Phe His Phe Leu Arg Val Tyr Asn  
 575 580 585

CAG CTG CGG GCC TCG CTG CAG GAC CTG AAG ACT GTG GTC ATC GCC AAG 3206  
 Gln Leu Arg Ala Ser Leu Gln Asp Leu Lys Thr Val Val Ile Ala Lys  
 590 595 600

ACC CCC GGG ACG GGA GGC AGC CCC CAG GGC TCC TTT GCG GAT GGC CAG 3254  
 Thr Pro Gly Thr Gly Gly Ser Pro Gln Gly Ser Phe Ala Asp Gly Gln  
 605 610 615

CCT GCC GAG CGC AGG GCC AGC AAT GAC CAG CGT CCC CAG GAG GTC CCA 3302  
 Pro Ala Glu Arg Arg Ala Ser Asn Asp Gln Arg Pro Gln Glu Val Pro  
 620 625 630 635

GCA GAG GCT CTG GCC CCG GCC CCA GTG GAA GTC CCA GCT CCA GCC CCG 3350  
 Ala Glu Ala Leu Ala Pro Ala Pro Val Glu Val Pro Ala Pro Ala Pro

640

645

650

GAA TTC GAT ATC AAG CTT ATC GAT ACC GTC GAC CTG CAG

3389

Glu Phe Asp Ile Lys Leu Ile Asp Thr Val Asp Leu Gln

655

660

664

## INFORMATION FOR SEQ ID NO: 2

## SEQUENCE CHARACTERISTICS:

LENGTH: 1954 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

## SEQUENCE DESCRIPTION: SEQ ID NO: 2

ATGACAGGCC AGGTGGGAGC GCAGACGGTC TCGGGTGGCA AAAGAAGCAT TGCAGGTCTG	60
ACACTTGTGA GGCCGCTCAG AAGTGTGCAC CTGCTTGATA TGTCCGTTCA AGTGATCAGG	120
CCTGGAGAAG CCTTTCCCAC AGCTCTGGCA GATGTAAGGC GGAATTCCCC AGAGAAGAAG	180
GGTGGTGAAG ACTCCCGGCT CTCAGCTGCC CCCTGCATCA GACCCAGCAG CTCCCCTCCC	240
ACTGTGGCTC CCGCATCTGC CTCCCTGCCC CAGCCCATCC TCTCTAACCA AGGAATCATG	300
TTCGTTCAGG AGGAGGCCCT GGCCAGCAGC CTCTCGTCCA CTGACAGTCT GACTCCCGAG	360
CACCAGCCCA TTGCCCAGGG ATGTTCTGAT TCCTTGGAGT CCATCCCTGC GGGACAGGCA	420
GCTTCCGATG ATTTAAGGGA CGTGCCAGGA GCTGTTGGTG GTGCAAGCCC AGAACATGCC	480
GAGCCGGAGG TCCAGGTGGT GCCGGGGTCT GGCCAGATCA TCTTCCTGCC CTTACCTGTC	540
ATTGGCTACA CGGCCACCAA TCAGGACTTC ATCCAGCGCC TGAGCACACT GATCCGGCAG	600
GCCATCGAGC GGCAGCTGCC TGCCTGGATC GAGGCTGCCA ACCAGCGGGA GGAGGGCCAG	660
GGTGAACAGG GCGAGGAGGA GGATGAGGAG GAGGAAGAAG AGGAGGACGT GGCTGAGAAC	720
CGCTACTTTG AAATGGGGCC CCCAGACGTG GAGGAGGAGG AGGGAGGAGG CCAGGGGGAG	780
GAAGAGGAGG AGGAAGAGGA GGATGAAGAG GCCGAGGAGG AGCGCCTGGC TCTGGAATGG	840
GCCCTGGGCG CGGACGAGGA CTTCTGCTG GAGCACATCC GCATCCTCAA GGTGCTGTGG	900
TGCTTCCTGA TCCATGTGCA GGGCAGTATC CGCCAGTTCG CCGCCTGCCT TGTGCTCACC	960
GACTTCGGCA TCGCAGTCTT CGAGATCCCG CACCAGGAGT CTCGGGGCAG CAGCCAGCAC	1020
ATCCTCTCCT CCCTGCGCTT TGTCTTTTGC TTCCCGCATG GCGACCTCAC CGAGTTTGGC	1080
TTCCTCATGC CGGAGCTGTG TCTGGTGCTC AAGGTACGGC ACAGTGAGAA CACGCTCTTC	1140
ATTATCTCGG ACGCCGCCAA CCTGCACGAG TTCCACGCGG ACCTGCGCTC ATGCTTTGCA	1200

```

CCCCAGCACA TGGCCATGCT GTGTAGCCCC ATCCTCTACG GCAGCCACAC CAGCCTGCAG 1260
GAGTTCCTGC GCCAGCTGCT CACCTTCTAC AAGGTGGCTG GCGGCTGCCA GGAGCGCAGC 1320
CAGGGCTGCT TCCCCGTCTA CCTGGTCTAC AGTGACAAGC GCATGGTGCA GACGGCCGCC 1380
GGGGACTACT CAGGCAACAT CGAGTGGGCC AGCTGCACAC TCTGTTGAGC CGTGCGGCGC 1440
TCCTGCTGCG CGCCCTCTGA GGCCGTCAAG TCCGCCGCCA TCCCCTACTG GCTGTTGCTC 1500
ACGCCCCAGC ACCTCAACGT CATCAAGGCC GACTTCAACC CCATGCCCAA CCGTGGCACC 1560
CACAACGTGC GCAACCGCAA CAGCTTCAAG CTCAGCCGTG TGCCGCTCTC CACCGTGCTG 1620
CTGGACCCCA CACGCAGCTG TACCCAGCCT CGGGGCGCCT TTGCTGATGG CCACGTGCTA 1680
GAGCTGCTCG TGGGGTACCG CTTTGTCACT GCCATCTTCG TGCTGCCCCA CGAGAAGTTC 1740
CACTTCCTGC GCGTCTACAA CCAGCTGCGG GCCTCGCTGC AGGACCTGAA GACTGTGGTC 1800
ATCGCCAAGA CCCCCGGGAC GGGAGGCAGC CCCCAGGGCT CCTTTGCGGA TGGCCAGCCT 1860
GCCGAGCGCA GGGCCAGCAA TGACCAGCGT CCCCAGGAGG TCCCAGCAGA GGCTCTGGCC 1920
CCGGCCCCAG TGGAAGTCCC AGCTCCAGCC CCGG 1954

```

## INFORMATION FOR SEQ ID NO: 3

## SEQUENCE CHARACTERISTICS:

LENGTH: 3318 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

## SEQUENCE DESCRIPTION: SEQ ID NO: 3

```

AATTCCAGTT TAATACTAAC CCTAATGTGT GACTGCGGTT TACAAAGAGC TCTGTATCAC 60
CTGGGATAGC TTTCAGTAGC AATTCACCTAC AACTGGTCCT AAAAAATAAT AACAATAATA 120
ATAATAATTA GAGAATTAAA ACCCAACAGC ATGTTGAATG GTTAAAATCA CGTAAGAACT 180
GAAATTTGGG GTGGGGGTGT CCTCAACAGC TGAGCTTGTC CTAGCAGTGA AAATGCTCGC 240
CTCCAAGCAG GGCTCAGAAA GGTCTGGAGC CCTCCAGGCA GAGGGCTGAG CTCAGGGGGC 300
TCTTGGAGGA CACTCACCCC ATGGTCCATG GGATGCTTCT GGCTTCCTTA AAAACAGTTG 360
GGCATCCGCA TTGTATAAGT AGGTGGAGAC CCTAGTGTGG TTCTTTTGAA GGATATGGGA 420
AGGGAGGATG ACGAACTAGA GAAGTGGGAG GGGACCAAAA TCACTGAGGT CCCAGAATAT 480
CATAGATTTG GGTATAGGAT TGGGGTCACT AAGAATTGAG CACCAGGAAT TCCAGCTTCT 540
TCCCATTAAA GAAACTGGGA CTGGTTTTGC CTTGGAGGCC TATGTAGTGT TTTCTGCCCC 600
TGTCCCATAC CAAGTCTCAT TGATATTTCT GCAGAATATC AGATGAAAAT CTATTTCTAA 660

```

AGACCATTGG	GAGAATGGGT	GGTGGAGAAG	GAGTTGGAGT	GGGGTTGGGG	GGCAGTTAAA	720
AATGAATAAA	AATCTCTCAG	CTACAGAACC	CAAACATCAC	TTCCCTCCGC	ATTCACAGCA	780
TTTCCCAGCA	GTCCCCAGAT	GGTTGTTTCC	GTGGGGACAC	AGCAGCTGCC	TCATTTCCCT	840
TCAGGCCCCA	TGGGCTGCTG	GTCAACCTCA	GGATCTACTA	AAGATGACGC	AAATGCCGAC	900
TGAACAATCT	GAAACCCAAA	GGACTCGAGG	AGAGACATGT	TCTGCTGAGG	AGAGAAAGGT	960
GAGCCAAGGG	CAGGGCCCAG	GTCCCCCAGG	GGGCCCCCGA	GAGCCCGGAC	ATGCACCTTC	1020
TGGATGTGTT	TGTTCAAGTA	GGACTTAGAG	CGGAAGAAGC	TCCCACATTC	AGGGCATGGG	1080
TACTTCTTCT	CCCCATCAGA	CTCCATTTTG	TTTTTGGGGA	CTGCCATGTC	GCAGGAGAAA	1140
GAGCCATTGG	CACTCTGCTT	CTCTGGCGTC	TTCAGGTCGC	TGGCATCTGA	GAGGTCACCA	1200
TAGGAGTCAG	AGCTCTCAAT	CGGATCCTGA	TGTGAGCATT	TCTGGCCTTC	TCGGTTACAG	1260
ATACTGCAGA	AGTTGCTGGG	CCCCTCGCTG	TGCTTCTTCA	GGTGGTCTGC	CATGTATGCT	1320
GCCCGCAAGT	ACTTCCCACA	CACCTGGCAG	GGCACCTTGT	CTTCATGACA	GGCCAGGTGG	1380
GAGCGCAGAC	GGTCTCGGGT	GGCAAAAGAA	GCATTGCAGG	TCTGACACTT	GTGAGGCCGC	1440
TCAGAAGTGT	GCACCTGCTT	GATATGTCCG	TTCAAGTGAT	CAGGCC'TGGA	GAAGCCTTTC	1500
CCACAGCTCT	GGCAGATGTA	AGGCGGAATT	CCCCAGAGAA	GAAGGGTGGT	GAAGACTCCC	1560
GGCTCTCAGC	TGCCCCCTGC	ATCAGACCCA	GCAGCTCCCC	TCCCCTGTG	GCTCCCGCAT	1620
CTGCCTCCCT	GCCCCAGCCC	ATCCTCTCTA	ACCAAGGAAT	CATGTTTCGT	CAGGAGGAGG	1680
CCCTGGCCAG	CAGCCTCTCG	TCCACTGACA	GTCTGACTCC	CGAGCACCAG	CCCATTGCCC	1740
AGGGATGTTT	TGATTCCCTG	GAGTCCATCC	CTGCGGGACA	GGCAGCTTCC	GATGATTTAA	1800
GGGACGTGCC	AGGAGCTGTT	GGTGGTGCAA	GCCCAGAACA	TGCCGAGCCG	GAGGTCCAGG	1860
TGGTGCCGGG	GTCTGGCCAG	ATCATCTTCC	TGCCCTTCAC	CTGCATTGGC	TACACGGCCA	1920
CCAATCAGGA	CTTCATCCAG	CGCCTGAGCA	CACTGATCCG	GCAGGCCATC	GAGCGGCAGC	1980
TGCCTGCCTG	GATCGAGGCT	GCCAACCAGC	GGGAGGAGGG	CCAGGGTGAA	CAGGGCGAGG	2040
AGGAGGATGA	GGAGGAGGAA	GAAGAGGAGG	ACGTGGCTGA	GAACCGCTAC	TTTGAAATGG	2100
GGCCCCCAGA	CGTGGAGGAG	GAGGAGGGAG	GAGGCCAGGG	GGAGGAAGAG	GAGGAGGAAG	2160
AGGAGGATGA	AGAGGCCGAG	GAGGAGCGCC	TGGCTCTGGA	ATGGGCCCTG	GGCGCGGACG	2220
AGGACTTCCT	GCTGGAGCAC	ATCCGCATCC	TCAAGGTGCT	GTGGTGCTTC	CTGATCCATG	2280
TGCAGGGCAG	TATCCGCCAG	TTCCGCCCCT	GCCTTGCTGCT	CACCGACTTC	GGCATCGCAG	2340
TCTTCGAGAT	CCCGCACCAG	GAGTCTCGGG	GCAGCAGCCA	GCACATCCTC	TCCTCCCTGC	2400
GCTTTGTCTT	TTGCTTCCCG	CATGGCGACC	TCACCGAGTT	TGGCTTCCTC	ATGCCGGAGC	2460
TGTGTCTGGT	GCTCAAGGTA	CGGCACAGTG	AGAACACGCT	CTTCATTATC	TCGGACGCCG	2520
CCAACCTGCA	CGAGTTCCAC	GCGGACCTGC	GCTCATGCTT	TGCACCCAG	CACATGGCCA	2580
TGCTGTGTAG	CCCCATCCTC	TACGGCAGCC	ACACCAGCCT	GCAGGAGTTC	CTGCGCCAGC	2640
TGCTCACCTT	CTACAAGGTG	GCTGGCGGCT	GCCAGGAGCG	CAGCCAGGGC	TGCTTCCCCG	2700

TCTACCTGGT	CTACAGTGAC	AAGCGCATGG	TGCAGACGGC	CGCCGGGGAC	TACTCAGGCA	2760
ACATCGAGTG	GGCCAGCTGC	ACACTCTGTT	CAGCCGTGCG	GCGCTCCTGC	TGCGCGCCCT	2820
CTGAGGCCGT	CAAGTCCGCC	GCCATCCCCT	ACTGGCTGTT	GCTCACGCCC	CAGCACCTCA	2880
ACGTCATCAA	GGCCGACTTC	AACCCCATGC	CCAACCGTGG	CACCCACAAC	TGTCGCAACC	2940
GCAACAGCTT	CAAGCTCAGC	CGTGTGCCGC	TCTCCACCGT	GCTGCTGGAC	CCCACACGCA	3000
GCTGTACCCA	GCCTCGGGGC	GCCTTTGCTG	ATGGCCACGT	GCTAGAGCTG	CTCGTGGGGT	3060
ACCGCTTTGT	CACTGCCATC	TTCGTGCTGC	CCCACGAGAA	GTTCCACTTC	CTGCGCGTCT	3120
ACAACCAGCT	GCGGGCCTCG	CTGCAGGACC	TGAAGACTGT	GGTCATCGCC	AAGACCCCCG	3180
GGACGGGAGG	CAGCCCCCAG	GGCTCCTTTG	CGGATGGCCA	GCCTGCCGAG	CGCAGGGCCA	3240
GCAATGACCA	GCGTCCCCAG	GAGGTCCCAG	CAGAGGCTCT	GGCCCCGGCC	CCAGTGGAAG	3300
TCCCAGCTCC	AGCCCCGG					3318

## INFORMATION FOR SEQ ID NO: 4

## SEQUENCE CHARACTERISTICS:

LENGTH: 1171 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

## SEQUENCE DESCRIPTION: SEQ ID NO: 4

GAGGAGGAGG	AAGAGGAGGA	TGAAGAGGCC	GAGGAGGAGC	GCCTGGCTCT	GGAATGGGCC	60
CTGGGCGCGG	ACGAGGACTT	CCTGCTGGAG	CACATCCGCA	TCCTCAAGGT	GCTGTGGTGC	120
TTCCTGATCC	ATGTGCAGGG	CAGTATCCGC	CAGTTCGCCG	CCTGCCTTGT	GCTCACCGAC	180
TTCCGCATCG	CAGTCTTCGA	GATCCCGCAC	CAGGAGTCTC	GGGGCAGCAG	CCAGCACATC	240
CTCTCCTCCC	TGCGCTTTGT	CTTTTGCTTC	CCGCATGGCG	ACCTCACCGA	GTTTGGCTTC	300
CTCATGCCGG	AGCTGTGTCT	GGTGCTCAAG	GTACGGCACA	GTGAGAACAC	GCTCTTCATT	360
ATCTCGGACG	CCGCCAACCT	GCACGAGTTC	CACGCGGACC	TGCGCTCATG	CTTTGCACCC	420
CAGCACATGG	CCATGCTGTG	TAGCCCCATC	CTCTACGGCA	GCCACACCAG	CCTGCAGGAG	480
TTCCTGCGCC	AGCTGCTCAC	CTTCTACAAG	GTGGCTGGCG	GCTGCCAGGA	GCGCAGCCAG	540
GGCTGCTTCC	CCGTCTACCT	GGTCTACAGT	GACAAGCGCA	TGGTGCAGAC	GGCCGCCGGG	600
GACTACTCAG	GCAACATCGA	GTGGGCCAGC	TGCACACTCT	GTTCAGCCGT	GCGGCGCTCC	660
TGCTGCGCGC	CCTCTGAGGC	CGTCAAGTCC	GCCGCCATCC	CCTACTGGCT	GTTGCTCACG	720
CCCCAGCACC	TCAACGTCAT	CAAGGCCGAC	TTCAACCCCA	TGCCCAACCG	TGGCACCCAC	780



```

AACTGTCGCA ACCGCAACAG CTTCAAGCTC AGCCGTGTGC CGCTCTCCAC CGTGCTGCTG      840
GACCCACAC GCAGCTGTAC CCAGCCTCGG GCGCCTTTG CTGATGGCCA CGTGCTAGAG      900
CTGCTCGTGG GGTACCGCTT TGTCACGTGC ATCTTCGTGC TGCCCCACGA GAAGTTCCAC      960
TTCCTGCGCG TCTACAACCA GCTGCGGGCC TCGCTGCAGG ACCTGAAGAC TGTGGTCATC     1020
GCCAAGACCC CCGGGACGGG AGGCAGCCCC CAGGGCTCCT TTGCGGATGG CCAGCCTGCC     1080
GAGCGCAGGG CCAGCAATGA CCAGCGTCCC CAGGAGGTCC CAGCAGAGGC TCTGGCCCCG     1140
GCCCCAGTGG AAGTCCCAGC TCCAGCCCCG G                                     1171

```

## INFORMATION FOR SEQ ID NO: 5

## SEQUENCE CHARACTERISTICS:

LENGTH: 651 amino acids

TYPE: polypeptide

STRANDEDNESS: single

TOPOLOGY: linear

## SEQUENCE DESCRIPTION: SEQ ID NO: 5

Met Thr Gly Gln Val Gly Ala Gln Thr Val Ser

1 5 10

Gly Gly Lys Arg Ser Ile Ala Gly Leu Thr Leu Val Arg Pro Leu Arg

15 20 25

Ser Val His Leu Leu Asp Met Ser Val Gln Val Ile Arg Pro Gly Glu

30 35 40

Ala Phe Pro Thr Ala Leu Ala Asp Val Arg Trp Asn Ser Pro Glu Lys

45 50 55

Lys Gly Gly Glu Asp Ser Trp Leu Ser Ala Ala Pro Cys Ile Arg Pro

60 65 70 75

Ser Ser Ser Pro Pro Thr Val Ala Pro Ala Ser Ala Ser Leu Pro Gln

80 85 90

Pro Ile Leu Ser Asn Gln Gly Ile Met Phe Val Gln Glu Glu Ala Leu

95 100 105

Ala Ser Ser Leu Ser Ser Thr Asp Ser Leu Thr Pro Glu His Gln Pro

110 115 120

Ile Ala Gln Gly Cys Ser Asp Ser Leu Glu Ser Ile Pro Ala Gly Gln

125	130	135	
Ala Ala Ser Asp Asp Leu Arg Asp Val Pro Gly Ala Val Gly Gly Ala			
140	145	150	155
Ser Pro Glu His Ala Glu Pro Glu Val Gln Val Val Pro Gly Ser Gly			
	160	165	170
Gln Ile Ile Phe Leu Pro Phe Thr Cys Ile Gly Tyr Thr Ala Thr Asn			
	175	180	185
Gln Asp Phe Ile Gln Arg Leu Ser Thr Leu Ile Trp Gln Ala Ile Glu			
	190	195	200
Trp Gln Leu Pro Ala Trp Ile Glu Ala Ala Asn Gln Trp Glu Glu Gly			
	205	210	215
Gln Gly Glu Gln Gly Glu Glu Glu Asp Glu Glu Glu Glu Glu Glu			
220	225	230	235
Asp Val Ala Glu Asn Arg Tyr Phe Glu Met Gly Pro Pro Asp Val Glu			
	240	245	250
Glu Glu Glu Gly Gly Gly Gln Gly Glu Glu Glu Glu Glu Glu Glu			
	255	260	265
Asp Glu Glu Ala Glu Glu Glu Arg Leu Ala Leu Glu Trp Ala Leu Gly			
	270	275	280
Ala Asp Glu Asp Phe Leu Leu Glu His Ile Arg Ile Leu Lys Val Leu			
	285	290	295
Trp Cys Phe Leu Ile His Val Gln Gly Ser Ile Arg Gln Phe Ala Ala			
300	305	310	315
Cys Leu Val Leu Thr Asp Phe Gly Ile Ala Val Phe Glu Ile Pro His			
	320	325	330
Gln Glu Ser Trp Gly Ser Ser Gln His Ile Leu Ser Ser Leu Arg Phe			
	335	340	345
Val Phe Cys Phe Pro His Gly Asp Leu Thr Glu Phe Gly Phe Leu Met			
	350	355	360
Pro Glu Leu Cys Leu Val Leu Lys Val Arg His Ser Glu Asn Thr Leu			
	365	370	375
Phe Ile Ile Ser Asp Ala Ala Asn Leu His Glu Phe His Ala Asp Leu			
380	385	390	395
Arg Ser Cys Phe Ala Pro Gln His Met Ala Met Leu Cys Ser Pro Ile			

	400	405	410
Leu Tyr Gly Ser His Thr Ser Leu Gln Glu Phe Leu Arg Gln Leu Leu			
	415	420	425
Thr Phe Tyr Lys Val Ala Gly Gly Cys Gln Glu Arg Ser Gln Gly Cys			
	430	435	440
Phe Pro Val Tyr Leu Val Tyr Ser Asp Lys Arg Met Val Gln Thr Ala			
	445	450	455
Ala Gly Asp Tyr Ser Gly Asn Ile Glu Trp Ala Ser Cys Thr Leu Cys			
460	465	470	475
Ser Ala Val Arg Arg Ser Cys Cys Ala Pro Ser Glu Ala Val Lys Ser			
	480	485	490
Ala Ala Ile Pro Tyr Trp Leu Leu Leu Thr Pro Gln His Leu Asn Val			
	495	500	505
Ile Lys Ala Asp Phe Asn Pro Met Pro Asn Arg Gly Thr His Asn Cys			
	510	515	520
Arg Asn Arg Asn Ser Phe Lys Leu Ser Arg Val Pro Leu Ser Thr Val			
	525	530	535
Leu Leu Asp Pro Thr Arg Ser Cys Thr Gln Pro Arg Gly Ala Phe Ala			
540	545	550	555
Asp Gly His Val Leu Glu Leu Leu Val Gly Tyr Arg Phe Val Thr Ala			
	560	565	570
Ile Phe Val Leu Pro His Glu Lys Phe His Phe Leu Arg Val Tyr Asn			
	575	580	585
Gln Leu Arg Ala Ser Leu Gln Asp Leu Lys Thr Val Val Ile Ala Lys			
	590	595	600
Thr Pro Gly Thr Gly Gly Ser Pro Gln Gly Ser Phe Ala Asp Gly Gln			
	605	610	615
Pro Ala Glu Arg Arg Ala Ser Asn Asp Gln Arg Pro Gln Glu Val Pro			
620	625	630	635
Ala Glu Ala Leu Ala Pro Ala Pro Val Glu Val Pro Ala Pro Ala Pro			
	640	645	650

INFORMATION FOR SEQ ID NO: 6

SEQUENCE CHARACTERISTICS:

LENGTH: 390 amino acids

TYPE: polypeptide

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 6

Glu Glu Glu Glu Glu Glu

1

5

Asp Glu Glu Ala Glu Glu Glu Arg Leu Ala Leu Glu Trp Ala Leu Gly

10

15

20

Ala Asp Glu Asp Phe Leu Leu Glu His Ile Arg Ile Leu Lys Val Leu

25

30

35

Trp Cys Phe Leu Ile His Val Gln Gly Ser Ile Arg Gln Phe Ala Ala

40

45

50

Cys Leu Val Leu Thr Asp Phe Gly Ile Ala Val Phe Glu Ile Pro His

55

60

65

70

Gln Glu Ser Trp Gly Ser Ser Gln His Ile Leu Ser Ser Leu Arg Phe

75

80

85

Val Phe Cys Phe Pro His Gly Asp Leu Thr Glu Phe Gly Phe Leu Met

90

95

100

Pro Glu Leu Cys Leu Val Leu Lys Val Arg His Ser Glu Asn Thr Leu

105

110

115

Phe Ile Ile Ser Asp Ala Ala Asn Leu His Glu Phe His Ala Asp Leu

120

125

130

Arg Ser Cys Phe Ala Pro Gln His Met Ala Met Leu Cys Ser Pro Ile

135

140

145

150

Leu Tyr Gly Ser His Thr Ser Leu Gln Glu Phe Leu Arg Gln Leu Leu

155

160

165

Thr Phe Tyr Lys Val Ala Gly Gly Cys Gln Glu Arg Ser Gln Gly Cys

170

175

180

Phe Pro Val Tyr Leu Val Tyr Ser Asp Lys Arg Met Val Gln Thr Ala

185

190

195

Ala Gly Asp Tyr Ser Gly Asn Ile Glu Trp Ala Ser Cys Thr Leu Cys

200

205

210

Ser Ala Val Arg Arg Ser Cys Cys Ala Pro Ser Glu Ala Val Lys Ser  
 215                                      220                                      225                                      230  
 Ala Ala Ile Pro Tyr Trp Leu Leu Leu Thr Pro Gln His Leu Asn Val  
    235                                      240                                      245  
 Ile Lys Ala Asp Phe Asn Pro Met Pro Asn Arg Gly Thr His Asn Cys  
    250                                      255                                      260  
 Arg Asn Arg Asn Ser Phe Lys Leu Ser Arg Val Pro Leu Ser Thr Val  
    265                                      270                                      275  
 Leu Leu Asp Pro Thr Arg Ser Cys Thr Gln Pro Arg Gly Ala Phe Ala  
    280                                      285                                      290  
 Asp Gly His Val Leu Glu Leu Leu Val Gly Tyr Arg Phe Val Thr Ala  
 295                                      300                                      305                                      310  
 Ile Phe Val Leu Pro His Glu Lys Phe His Phe Leu Arg Val Tyr Asn  
    315                                      320                                      325  
 Gln Leu Arg Ala Ser Leu Gln Asp Leu Lys Thr Val Val Ile Ala Lys  
    330                                      335                                      340  
 Thr Pro Gly Thr Gly Gly Ser Pro Gln Gly Ser Phe Ala Asp Gly Gln  
    345                                      350                                      355  
 Pro Ala Glu Arg Arg Ala Ser Asn Asp Gln Arg Pro Gln Glu Val Pro  
    360                                      365                                      370  
 Ala Glu Ala Leu Ala Pro Ala Pro Val Glu Val Pro Ala Pro Ala Pro  
 375                                      380                                      385                                      390

## INFORMATION FOR SEQ ID NO: 7

## SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

## SEQUENCE DESCRIPTION: SEQ ID NO: 7

CTTGAGGATG CGGATGTGCT 20

## INFORMATION FOR SEQ ID NO: 8

## SEQUENCE CHARACTERISTICS:

LENGTH: 18 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 8

CCATGGGGTG AGTGTCTT 18

## INFORMATION FOR SEQ ID NO: 9

## SEQUENCE CHARACTERISTICS:

LENGTH: 18 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 9

AGGACACTCA CCCCATGG 18

## INFORMATION FOR SEQ ID NO: 10

## SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 10

GTATGGGACA GGGGCAGAAA 20

## INFORMATION FOR SEQ ID NO: 11

## SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 11

TTTCTAAAGA CCATTGGGAG 20

INFORMATION FOR SEQ ID NO: 12

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 12

CCATTTTAAA GTAGCGGTTC 20

INFORMATION FOR SEQ ID NO: 13

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 13

AGGAGAGAAA GGTGAGCCAA 20

INFORMATION FOR SEQ ID NO: 14

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 14

GTAGATCCTG AGGTTGACCA 20

## INFORMATION FOR SEQ ID NO: 15

## SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 15

TGTGAGCATT TCTGGCCTTC 20

## INFORMATION FOR SEQ ID NO: 16

## SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 16

TGAAGACGCC AGAGAAGCAG 20

## INFORMATION FOR SEQ ID NO: 17

## SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 17

GCCTCACAAG TGTCAGACCT 20

## INFORMATION FOR SEQ ID NO: 18

## SEQUENCE CHARACTERISTICS:

LENGTH: 18 base pairs

TYPE: nucleic acid



STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 18

AGAAGGGTGG TGAAGACT 18

INFORMATION FOR SEQ ID NO: 19

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 19

CTTGGTAGA GAGGATGGGC 20

INFORMATION FOR SEQ ID NO: 20

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 20

GCCCATCCTC TCTAACCAAG 20

INFORMATION FOR SEQ ID NO: 21

SEQUENCE CHARACTERISTICS:

LENGTH: 15202 nucleic acids

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

FEATURE:

NAME/KEY:

## LOCATION:

## IDENTIFICATION METHOD:

OTHER INFORMATION: /note="N is unknown or other"

SEQUENCE DESCRIPTION: SEQ ID NO: 21

GATCCGAGCT	CAATTAACCC	TCATAAAGG	GAGTCGACTC	GATCCTTAAA	ATATTCATAT	60
CTCCTGGACA	ACCTGTGGCC	ATAGTGCCTG	ACTGTAAACC	CAAAGGGTTT	GCCTTTGCCA	120
GTGTAGCCCA	GCCTGGTGTC	TGCTGCCCCT	CGCGGTGTCT	GTGCACCTGC	CACGATGCTG	180
ACCAGACACC	CTTAACCAGG	TTCACCCATC	GCCTGGGCCT	GGAGCAGTCC	CCCTGATGCT	240
CTGATTGGTC	CTTGGACCTT	CTGTTCTCCC	AAAATCCAG	GTCAGAAAAT	ACCTGGAAGT	300
CTATTTGTGT	CCCACCTCCC	TCTTTGTGGC	CGCAAGTGCC	CCTTCCTCCA	CACAGTCACA	360
AGACCATGAG	ATGCCATCTC	CTCCCCCTCT	GGGCTGCAGA	CTTTGGGAAG	CTCCCAGGCC	420
ACAGAGGTGT	CAGCTCCTGT	CCAGGCCCTT	GGGACCTTCC	CTCATTCAAC	CACCCCTACCC	480
AACCCCCCAC	TGCCTGCCAG	CCACCACTCC	CTCCCACATT	TGCAGGCGGG	GGCCCTGCCC	540
TCTCCTGCCG	CTGGTTCCCC	TACCCAGGAG	GCTCTCCCAT	CGCTCTTTTG	AGAGTCTGCC	600
TCCCACCTCT	AACTGGGGGC	TTAGTTCAAG	TTGCCCCCTT	ACCCTAGTCC	CAGCTGCCCA	660
AGAGCTTGCT	GCCTCCTGTT	CTTGGTGAGG	GACTCCAGAG	ACAGATGTGA	GACCTCCCTG	720
GACCCCTCCA	AGGCATTCCC	AGGTCACCTC	CATGAGTAGT	GAAGAACCGC	CTCTGAGCAG	780
GCTGAGCCTC	CCTCAGCCTA	TGGTGTCTCT	ACGTGGCTTG	GCCCACAGCA	GGTGCTCACG	840
CCTCCTCCTC	AGCAGAGCCT	ACCATCCTCC	TGCCATGCTC	ACCAGTCCCC	ATGCTGATAG	900
CCATCACCAG	TCCCCATGCT	GATAGCCATC	ACCAGTCCCC	ATGCTGATAG	CCACTTTCTG	960
GATGCTCTAG	GTCTGTCTGG	ATGACACAGT	GACCACAGAG	AAGGAGCTGG	ACACTGTGGA	1020
AGTGCTGAAA	GCAATTCAGA	AAGCCAAGGA	GGTCAAGTCC	AAACTGAGCA	ACCCAGAGAA	1080
GAAGGTGGGT	TTGTGTGGCA	GGTGGGAGGG	CAGTGGTGCA	GAGCCAGCCG	GGATAGGAGC	1140
CAGTTGGGGG	GGCTTGGGCC	ATGGGACTGC	TCAGGGCTGC	CGAGTCCCAG	CTGCGCCCCT	1200
CCCTGGCTGC	ATGACCTCGG	GCAAGTCGCG	GCCTCTCTGT	TCTCTGTGGG	GTGGGGACAG	1260
TGGTAGTTCC	TGCTCTAAGG	ATATGATGAG	ACCATCTTTA	CCACCCAGTT	GGTGGGAACC	1320
GTTGCGCTCC	CTCCTCACAC	CCCTGGCCTT	GGGGAGCTCT	GTGCTTCCTC	TTCTCTCCCC	1380
GGCTGACTCA	AGCACTCGTC	CTCAGGGTGG	TGAAGACTCC	CGGCTCTCAG	CTGCCCCCTG	1440
CATCAGACCC	AGCAGCTCCC	CTCCCACTGT	GGCTCCCGCA	TCTGCCTCCC	TGCCCCAGCC	1500
CATCCTCTCT	AACCAAGGTA	ATCGTGTATG	TATCTTGCTT	CTAGTGGAGC	CACACAGCCC	1560
TGCCTGGGCC	CCCTGGCTGG	GCTGGGGTTG	GGGGAGAGGT	GCCAGCACCT	GCTTCCAACA	1620
GGGTCAGACA	CAGGGAGGGC	AGTGCCTTCT	GCAGGCTGGT	CCTCGCGGGG	GGACACATGG	1680
CAGGGGTGCC	TGGCCTGATG	CCAGCTGTTG	CTTGCTTGGT	GAGGACTCCC	AATTGCTCTG	1740

ATGCCCACAT	CCAGCTCCTC	TAGGAGACCG	CAGGGTGTCT	GACAGGCCCT	GAGGCTGCCC	1800
TCTGAACAGG	CTCGGGGCTG	TTGGCTCATG	GGACCCATTC	CCTCACCGGC	AGCACAAGCA	1860
GGTTGGCTCC	TGGTTACAGG	AAGCCGGGCT	TGTGACTTTA	CTGTCTGGAG	CCCGAATCCC	1920
TGTGCAGGGA	AAAGCTTGCT	TTTATCACTG	CCTCATCTCT	GTGGGGTGAC	CCAGCCCCAG	1980
AACACCATGT	TTGTGGGGCC	AAGATGGGCC	ATCTCTGTCC	CTGTGGACCC	ATGGAAGACC	2040
AGGCCCATTC	GTCTGCCCAC	TATCTTAGCG	TTTTCAAAGG	GCTTTCACCT	CTGAACCCAG	2100
GCATCCTCGG	AGATGAGTGA	GTGAAGCAGG	TCTCATGAGC	GTGTCTGCTG	GCCCGGCCCC	2160
CACGGAAGAG	GGGAGGGTGT	GCCGTCCCGA	GTGGAGCCGA	GGCTCGGGAC	ACGCAGGAAA	2220
GGACGCCGCC	TGCCCCGGGCT	CCTGGAGACG	CAGAACTTGG	TGTGAGGTCT	TGGGAAAACA	2280
GTTCAACCCG	ATGTTTTAAG	AGCCAGAAAA	ACATTCCCAC	CCCTTGACCT	GGTAACCCCA	2340
CTGGTGGGGA	TTTTCTCTTA	GAGGGATAAG	ATACCGGGAA	GGGGAGGTGA	AATGCTCACC	2400
ACTGCCAAAA	CACGGGCTGC	AACTGCAACA	TCGGAGGATG	AGAGGGAGAG	TCGGCTGTGG	2460
TGCAGAATGC	TCAGCAGCCC	TCCCAGCAGG	GACAGGAAGA	CTGGGCAGGA	AGAGGGGAGA	2520
AGCATTCAAG	TTAAGGCAAA	AGGCCCAACG	CAGAGCAGCA	CACTGAGGTC	ACACCTGTGA	2580
GATGTGGAAG	AGAATTCCTG	AGCGTGGAGC	GATGGGGTTA	GGTGCCAGGA	TGATTGCCCA	2640
TTTTGCTTCT	GTCAGACTCT	TGACTAAGGA	TTTCTGGTTG	CATTTTATTA	CATAAAAGCC	2700
AGGGAGGTTA	TATCACGGTG	AGAAAGCTTC	CCTGACGCCG	CCTCCTGTAG	CGCAGCCAAG	2760
CGAGCCTGTG	GAGGTACCAT	ATGACTGTAG	GCCTCTGGGG	ACAGGGAGCT	GCATCTGCTT	2820
CTCAAGGCCA	GGGACACAGC	CATTTCTGCC	AGCATCTGTT	GATCAGTGAG	TGAGTGAGTG	2880
GGCAGGTAGA	GCAGGAGCCA	GTGAAGAGCA	GGCCCTGGAT	GGGTGGGGAT	GCACCATGTC	2940
CCCAGGCTGC	AGCTGCAGGC	AGCCCCCCAC	ATTGTCGGAG	AAGCCTCTGC	ACCAGCTCAG	3000
CCCCCTCCTC	ACTCCCCTTG	TGCCCTGGGG	ACACTCTGCA	GAGGGGCACT	CTGCAGTCTG	3060
TCCCCGCCAT	CGCTGGACTT	CTGGACATGG	CCTCCAGATT	TGCACCTCTT	AAATAAATCT	3120
GCAGTGGATG	TCTTTGTGTG	CACCTCTCTT	TCCTTTTGGT	GAGAAACAGC	AAAGATCGGA	3180
CCCCTAAGGA	CTCTCCTGAT	GTCTCCGCTC	TATCCGCTGA	GTGCCCTTTC	TGACCACTTG	3240
TTTGTACAGG	CCACGGTCCA	GGACGGGAGC	AGATAGACTG	TCCCTGTCCC	TGTCCACATT	3300
TCCTTGGTCC	AAACAGGGCT	TGTGGGAGGT	AGTGGCAAAA	GGTGTTGGTC	TTTTTCTCAC	3360
TGATTTGGAG	GCCTCCCCGT	GTGTTTTTTC	AGCCGCGTGT	TCCTGGGTCT	TGCCTGGATG	3420
GACAGGGTTT	TTTAGCGCGT	GGGAGCAGCT	TTGCTGACCA	TGCCTGTTGC	TTCCAGCCTG	3480
ATTCCCGAGA	AGGGAGCGTG	CTTGCGAAGG	AACTGGCACT	CGGGCCTGCC	TGAAGGGGGC	3540
GCTGTCCAGA	CACACCCAGC	CTCCCGTCGT	GGCAGGCGCT	GTCGGAGCCA	TGGATGATTG	3600
TGACCAATAG	GGGTGGTCGC	CAGAGTTGAT	TGTCCAGCCA	GGCCCAGGGG	CTGAGAGGAG	3660
GCTGTGTGGA	GAGGTGGTTA	GGAGCCAGGG	CTCGGTCAGC	TGAGTTCGCA	TGCCAGCTTC	3720
CTAGCTGTGG	GACCTCAAGC	AACTTGTAGC	CCCTCTGAAG	CTGTTTTCTC	AACTGTGAAG	3780

TGGACGCACC	CTACTTCATT	GATTCTAAGA	GGCACGCATT	TCCACCTTGT	GACTTCTCTG	3840
AAACTGAGGT	GCGTCTTTCA	GTCAGTGGCG	TCTCATAGTC	GCTGTCAGCC	AGCTGGTATT	3900
CGAGATGGAG	TCGTGGAAAA	CCCGTGGACA	CCTTCCGCTA	GGACCAAGAT	GGCGCCACCT	3960
GCCGCATCTT	AGATTTGATG	AAATGTGGTA	AATAACGAGA	GGCATGCATG	AGCGAATGCT	4020
GGGGAGGCGC	TTGGCACTAC	CCAGAGCTCC	ACAGAGGTGG	TCGATGAGGG	CTGCCCTTTC	4080
CCACATCCTT	AGTAGGGGGT	TCAAGATGAC	CCAGACTGTG	CCCCTGGGGA	GCTTGGAGCC	4140
ATGCGGGAGG	ATGAGCCATG	TGCTGGAGGA	GAACAGGGTA	GGATGGTGTG	GGGCTTTTGT	4200
AGACTGTCTA	GAGCAGAGAA	GGTCTGCAGT	GGAGGTGGTG	TCTGAGGTGA	ATCTCGAAGG	4260
TGAATAGGAG	TTGAACGTTA	GCAGGCAGAG	GGTGGATTGC	AGGAGAGCAG	CGGCCTGGGC	4320
AGGTGCCCAG	CGTGGCCCAT	CAGGGTGCTT	CATGCATGGC	TGTGTGCTTG	CCATCCTTCC	4380
TGCCTGCCTA	CCCCCTGCTG	CTTCGCTTCA	TGGGGGCGTT	TGAGCTTGGG	CCCACCTGCC	4440
TGCCTCGCTT	GTGGGCAGAG	GACCCAGGCT	GTGTGAGTTG	TCCTGTCCCG	GGGAGCAGCT	4500
GAGCTTGTCC	GGGTTCTCG	ACCTGTGGGG	CTTCAGAGGA	CTTCGGGTCA	TTTCAATGGG	4560
CTGTGGCGAT	GCTGGCTGTG	GAGGTAGCCT	AGGGCTCCTG	TAGCCTTCAG	TGAGACTGGC	4620
GGCCCGATGC	CCAGTGTTCA	CCCTGCTGGC	GGCAGTCAGG	AACATGTTCA	CAAAGCTTTA	4680
CTTCAAGTGG	TCTAGAGGTG	ATCTGAGGTG	GAGTAACAGG	TCCAGATAGG	CTACGTTCAT	4740
AAAACAGCTT	CAGCGGGGTT	TAGGAACACT	GTGCATTTAC	GGGACGCAGT	GGGTCAGAGT	4800
GCTGCTGTCC	GTGGGAGGTG	GCCCCAGGGC	AGGTCAGTGG	GCACGTCCTG	TGGTAAGTGG	4860
GACTGTGGAT	GTGGGCTCAG	GCTGGACTCA	GCAGCCCTGC	TGGATACCAA	GGCCTGCAAG	4920
GGCTGGCCCC	CTGGTGAATT	GTCCCGTGCC	CTGTGTATCT	ATGAGTCCTG	CAGAGATGAC	4980
AAATCAGGGG	ACGGGGTCAT	GTCTAGTCAC	CGTCTGGGAA	AATGCTCCAG	GAGTGAACAC	5040
ATTTCAAGCT	CTTGATGGAT	GTACCTCCAA	ACTCTTCTCT	GGATGGGTGG	GCCAGCTTGC	5100
ATGCCTGTGC	CGGCCTCTGC	CCAGCGAGGT	CAGGGCCAGG	CCACACAGTC	AGTCTGACTT	5160
TGGCAGAAGT	TGAGAGGCAA	CACTTGTCTC	TTGTTTCAGC	TTGCCTTTCT	TTGTGTACTT	5220
CTGAGAGCGA	GCATTCTTTT	CATGTTCTAT	CCGCTGGCCG	TTCTTCTGCG	GAATGTCTGT	5280
TCACGTCCTT	TGCAGTCTGT	TAATGAGGTT	TCCAACCTTC	CCTCATTTTT	GTAATCTGTA	5340
AGAACTTTTT	CCAGACTAGC	GATATAAATC	CTTGTCAAAT	ATTGCAAACA	CTTTTCTCAT	5400
TTCATCTGGT	TTTAATCTAT	CCTGGTTTTT	AAAAAATGTG	TCTGTGGAAG	TTTAATTTTT	5460
ATGTAGTCAC	ATCTCAGTTT	TTTTCCATTG	CATTTATTCT	CAGAATGCTT	CTCCCTGCCC	5520
TGAGATTAGA	TAAGCAGTCA	TTTGTTCTTT	CTTGAGTTAT	TTTGAGATTT	CAGTTTTAAC	5580
ATTTTCTTCT	ATAATCCATG	TGGCTGGGTT	TTGGGATCTG	GCTAACCCCC	GCCATGCCAG	5640
TAGCCTGAGG	GGCCCAGCCC	CACTTGTTGA	ACAGCCGCTC	TCCCCGCCCC	ACCCACCCTG	5700
CCTGCCTGCC	CACCCGCCCT	GGTCTCTCCA	GGAATCATGT	TCGTTCAGGA	GGAGGCCCTG	5760
GCCAGCAGCC	TCTCGTCCAC	TGACAGTCTG	ACTCCCGAGC	ACCAGCCCAT	TGCCCAGGGA	5820

TGTTCTGATT	CCTTGGAGTC	CATCCCTGCG	GGACAGGTAA	TGCCCTCTTC	CCGCTTCTGG	5880
GGACCATACA	TCTGTGGGTG	GACTCTTCTG	CTTGGGGTTG	TGTGCAGTAG	GAAGTGGCCT	5940
AGCTGGAGCT	GAGGCAGATG	CTTCCAGGGT	TTGGCGTCTT	CTGCTTTGCG	CCACGGTCTT	6000
TCTCTTGGAC	CTGTCTCTGG	TTGAGTGTCT	TCCTGACAAA	CACAGTGGTT	AAGGGTTTAT	6060
TTTCAGCCTC	CCTCCTTCCC	TTCCCCACCC	ACCTTGGTTG	ATGGGAACAG	GCAGTTCTCT	6120
GTCACTGGGC	CCAGGGCAGC	AGGGGGGCAG	GTGGAGAGGG	TGGCCCTTGA	CCCTGTGAGC	6180
AGGCTTCCCT	GGGGAAGGCA	TTTCAAAAGA	CCCTCGTGCA	GGGGCTTGT	TGGGTTTCTT	6240
CTCTGTTTCC	TGGCACCCCT	GGAGCCACTC	GGCGCCTTTC	CGCATGTCAC	CCTGGTGGTC	6300
TGGGAAACAG	TCTCACTCTG	GCGCCTCCTC	TGTGGTTGTT	ACTGAGAGTT	CTGGGGCCCC	6360
TTCCTTTGTC	CTGAGGAAAG	ACAGGAGGAA	AGCAAGGGTG	CTTGCTGTGT	GCTTCGCAAA	6420
TGTGCTTGGT	GCCTGGGCCT	CCCTCCAGCC	CCATCTCTGC	AGCAGCACAA	GGTTATGGCC	6480
TTGTGACACT	GGGACAGTTT	GCAGAGTCCT	TGTCTGTCCT	CAGTACTCCA	CAGTATTCTG	6540
CCATCACCCCT	TTCCAGGGTC	ACACAGCAAG	AGATTCCCAA	GCCCTAGGTA	TTCCCCAGTG	6600
CACAGAGACC	ATTGGGAGGG	ACTTGCCAGG	GCTGTGTCCA	CTGCTGGCCA	GTTAGGGTCG	6660
GACCAAATTT	GTAGACTGTC	TACCTGGACC	CTTGCGTGGC	ACAAGGAGCA	GTCAGATGCT	6720
GGATCCCTGG	AGAGTGGCGA	GAGGCTCTGG	CCTTAGGTTG	CGAGTGGGAA	TCCCAGCCCT	6780
GCTGTGTGCT	GGTGGGATAA	CCAAGTGGGT	CTCTGCCCTT	GGGTCCCAGA	GTGGGCCCCA	6840
GGGTCCCAGA	GTGGGCTCCA	GGGTACAGCG	TGGGGATGGG	GAGCCTCCTC	AGGGCGGTGA	6900
TGGAGGGCAG	AATGCCCAGC	TCAGGGTCTG	GCAACCAGTA	AATGGCTGGG	GCTGGCTGCA	6960
GTAGGTGGGG	ACTGACTGTG	TTTCTTTCTC	CATCAGGCAG	CTTCCGATGA	TTTAAGGGAC	7020
GTGCCAGGAG	CTGTTGGTGG	TGCAAGGTAA	GGAAGAGGTT	GGAAAGGGAC	CTGGGCCTGG	7080
CCACACAGCC	TTATGCACAC	ACACTGCTGT	GGGCCAGGGG	TGGCCAGTCA	GGTTTTTTTA	7140
AAAATCCGTT	CACAGAAGGC	CTATAGAACT	ATTTCTTCCT	CTAAAGAGAC	ACAGATGAGA	7200
TGGACTTTTC	AATCTGTTTC	CAAATTCTAA	TACCTAAACT	CTGCTCAGCA	CATGTTGCCC	7260
TACACCAGGG	GTTGGCAAAT	CAAGGCCTGT	GTGTGGCCCA	CAGCCTGGGA	GCTAAGAATG	7320
ACAGTTACAT	TCTTTTTTCT	TTTTTTGAGA	CTGAGTCTCG	CTCTGTCGCC	CAGGCTGGAG	7380
TGCAGTGGCG	TGTTCTTGGC	TCACTGCAAC	CCCCGCCTCC	CAGATTAATG	CAATTTTCCT	7440
GTCTCAGCCT	CAGCCTTCTG	AGTAGCCCGG	ACCACAGGCG	CACGCCACCA	CGCCCAACTA	7500
ATTTTTTATA	TTTTTAGTAG	AGACAGAGAT	TCACCATGTG	GCCTAGCTGG	TCTCGAACTC	7560
CTGAACTCCA	GTGATCCACC	AACCTCGGCT	TCCTAAAGTA	CTGGAATTAC	AGGCATGAGC	7620
CACCGCGCCT	GGCTAGAATA	ACAGTTACTT	TTTTTTTCTT	TGAGACTGAG	TCTTGCTTTG	7680
TCACCCAGGC	TGGAGTGCAG	TGGCACGATC	TCAGCTCGCT	GCAACCTCCG	CCTCCCGGGT	7740
TCAAGCGATT	CTTCTGCCTC	AGCCACCCAA	GGTGCCCGCC	ACCACACCTG	GCTAATTTTT	7800
CTGTTTTTAG	TAGGGACAGG	ATTTCGCCAT	GTTGGACAGT	TACATTCTTA	AAGGGCTGCT	7860

GAAGATCGTA	TGGACATGGT	AGCCCATAAA	TCCCAAATG	TGTACTCTGA	CCCTTTACAG	7920
AAGCTTACTA	ACTCCCCTC	TACATGTGAG	GGCTGCGGTG	GCCAAGAAGA	GCTGGAATTT	7980
AAGTGTGAAG	GTCCTAAGAC	CTGCCCCAGC	CCACTTCCCT	GCCCCGGAGG	CCACCAGGGG	8040
TGACAAGTAG	ATTCATGCCC	TGGAGTGTTT	CTTCTCTCCG	GGGCTTATGG	CAGCAACTGA	8100
ATGACTTAGA	AGTCCATGGG	AGTGCTTTCT	GTTGTGGGAA	CTCGTGTGGT	CTGGGCATAG	8160
CTGTGCCAGG	CACCTATGGT	CCAAGCCCCT	AGAAGCATAG	ACTCTGACCA	AACTGGCGAC	8220
CCAGCCTTCC	AGCAGGCAGC	ACTGGCTCCC	ACCAGGGCCC	TCATCCTGGG	AACTGACTTG	8280
GCCATGTGGG	AGGCTTGGGA	GACCCATGGG	TTGGTTTCTC	AGGGTCAGGG	TGTAGCAGTG	8340
GGCTCCAGAT	GTGGCAGGTG	GGAGGTGGGA	GGGGCCCCTC	CCAGCATGCC	ACTGACCTGG	8400
CCTCTCCCTG	CACAGCCCAG	AACATGCCGA	GCCGGAGGTC	CAGGTGGTGC	CGGGGTCTGG	8460
CCAGATCATC	TTCCTGCCCT	TCACCTGCAT	TGGCTACACG	GCCACCAATC	AGGACTTCAT	8520
CCAGCGCCTG	AGCACACTGA	TCCGGCAGGC	CATCGAGCGG	CAGCTGCCTG	CCTGGATCGA	8580
GGCTGCCAAC	CAGCGGGAGG	AGGGCCAGGG	TGAACAGGGC	GAGGAGGAGG	ATGAGGAGGA	8640
GGAAGAAGAG	GAGGACGTGG	CTGAGAACCG	CTACTTTGAA	ATGGGGCCCC	CAGACGTGGA	8700
GGAGGAGGAG	GGAGGAGGCC	AGGGGGAGGA	AGAGGAGGAG	GAAGAGGAGG	ATGAAGAGGC	8760
CGAGGAGGAG	CGCCTGGCTC	TGGAATGGGC	CCTGGGCGCG	GACGAGGACT	TCCTGCTGGA	8820
GCACATCCGC	ATCCTCAAGG	TGCTGTGGTG	CTTCTGATC	CATGTGCAGG	GCAGTATCCG	8880
CCAGTTCGCC	GCCTGCCTTG	TGCTCACCGA	CTTCGGCATC	GCAGTCTTCG	AGATCCCGCA	8940
CCAGGAGTCT	CGGGGCAGCA	GCCAGCACAT	CCTCTCCTCC	CTGCGCTTTG	TCTTTTGCTT	9000
CCCGCATGGC	GACCTCACCG	AGTTTGGCTT	CCTCATGCCG	GAGCTGTGTC	TGGTGCTCAA	9060
GGTACGGCAC	AGTGAGAACA	CGCTCTTCAT	TATCTCGGAC	GCCGCCAACC	TGCACGAGTT	9120
CCACGCGGAC	CTGCGCTCAT	GCTTTGCACC	CCAGCACATG	GCCATGCTGT	GTAGCCCAT	9180
CCTCTACGGC	AGCCACACCA	GCCTGCAGGA	GTTCTGCGC	CAGCTGCTCA	CCTTCTACAA	9240
GGTGGCTGGC	GGCTGCCAGG	AGCGCAGCCA	GGGCTGCTTC	CCCGTCTACC	TGGTCTACAG	9300
TGACAAGCGC	ATGGTGACAG	CGGCCGCCGG	GGACTACTCA	GGCAACATCG	AGTGGGCCAG	9360
CTGCACACTC	TGTTACAGCCG	TGCGGCGCTC	CTGCTGCGCG	CCCTCTGAGG	CCGTCAAGTC	9420
CGCCGCCATC	CCCTACTGGC	TGTTGCTCAC	GCCCCAGCAC	CTCAACGTCA	TCAAGGCCGA	9480
CTTCAACCCC	ATGCCCAACC	GTGGCACCCA	CAACTGTGCG	AACCGCAACA	GCTTCAAGCT	9540
CAGCCGTGTG	CCGCTCTCCA	CCGTGCTGCT	GGACCCACCA	CGCAGCTGTA	CCCAGCCTCG	9600
GGGCGCCTTT	GCTGATGGCC	ACGTGCTAGA	GCTGCTCGTG	GGGTACCGCT	TTGTCACTGC	9660
CATCTTCGTG	CTGCCCCACG	AGAAGTTCCA	CTTCTGCGC	GTCTACAACC	AGCTGCGGGC	9720
CTCGCTGCAG	GACCTGAAGA	CTGTGGTCAT	CGCCAAGACC	CCCGGGACGG	GAGGCAGCCC	9780
CCAGGGCTCC	TTTGCGGATG	GCCAGCCTGC	CGAGCGCAGG	GCCAGGTGAG	ATCAAGCACA	9840
GCTCTCAGGG	GCCCCGGGGG	CACGGGTCTG	GCATGTGTGT	GATCTCAGCA	TCTGCGGCTA	9900

GTGTGGGCTG GGAGTTGCTG CGAGAGCTGG GCCCCCTCCC CCCTGCCCCCT CGCCCCCCCC 9960  
GGGCCTCCCT CTACATCACC ACCCCAGGTT TGGTGCCAGG CTGCTCCTTA TCTCAGTGCT 10020  
GTAGAAGAAG CCCAGGAAAG CTGTCCTCTC ACAAATGGG TTGGCCCAGC CTCTTGCCAC 10080  
CCATGAAGGG CAGGCCAAGG GGGCTGCCCC ACCTTTGCCT GCCCAGTGGG AGAGCAACAG 10140  
GCTGCAGCAC ACCGAGGCCA GGAGAGCTGT CACCCTGGCT GCTGTGCTCC TCTGGGCCCA 10200  
AGCATGGCCT CTGGGCACTA CCTCCTCCAG GGTCACAGTC CCACGGATGG CTCTGTGGGC 10260  
CAGGATCTGC CTTAGGCTTC ACCCACCTCA ACATCTTGCT GTGTTGTTCA GGCTGGTCTC 10320  
AAACTTTGGG CTCAAACAAT CCTCCGCCTC AGCCTCCCAA AGTGCTGGGA TTACAGACAT 10380  
GAGCCACCGT GCCCGGCCGT GCTGTTCTGT TCTCCAATAG AGAAGCTGGT GGAAGTCCCC 10440  
AGTAACCCAG AGGTGATGTG TGATGCACAC AGTCTCCTCA CTCTGAAGCT GCACATGCGA 10500  
TGTGAATCTT CATTTGGGGT CCGCTGTTAA TATGGTGTTC TTCGGGGGAT ACAGCAATGA 10560  
CCAGCGTCCC CAGGAGGTCC CAGCAGAGGC TCTGGCCCCG GCCCCAGTGG AAGTCCCAGC 10620  
TCCAGCCCCCT GCAGCAGCCT CAGCCTCAGG CCCAGCGAAG ACTCCGGCCC CAGCAGAGGC 10680  
CTCAACTTCA GCTTTGGTCC CAGAGGAGAC GCCAGTGGA GCTCCAGCCC CACCCCCAGC 10740  
CGAGGCCCCCT GCCCAGTACC CGAGTGAGCA CCTCATCCAG GCCACCTCGG AGGAGAATCA 10800  
GATCCCCCTG CACTTGCCCTG CCTGCCCCGTC GCTCCGGCAC GTCGCCAGCC TCGGGGGCAG 10860  
CGCCATCATC GAGCTCTTCC ACAGCAGCAT TGCTGAGGTA GCGGCCCCGG TGTGGGTGCC 10920  
AGCTATGGCA CGGCCAGTCC TGAGGGCGAG GCCAAGCTTG GCTTCAGGTC AGCCTCAGGT 10980  
CCCTGGACTT CCCTGATGTC GGAGTCCTCA GCTGAGCTGC TCACAGCTTT GAGGACCTGG 11040  
GCAGTGAGGT CCTGAGTTGC CCTCCCTGGC CATTTGTGCT GTGTCACCAC CTCCTGTGCC 11100  
ACTTCCAGCC CCAGGTAGAC CTCCCACCAA CAGCCATCTC CCACCCCTCT CTCCTCTCT 11160  
GCCTTGAAGC ATACGGATTG ATTGGTGAGC CAAGAGGGGC TTCCCATGTC TCCTTGTGGA 11220  
AGCTGTGGGC ATGTCCTTGG TATGTGCAGG TTGCTAGGGT GGTGGAGCTG ACAGGAGGCC 11280  
CCCCGTCTTC AGGTTGAAA CGAGGAGCTG AGGCACCTCA TGTGGTCCTC GGTGGTGTTC 11340  
TACCAGACCC CAGGGCTGGA GGTGACTGCC TCGTGCTGC TCTCCACCAA GGCTGTGTAC 11400  
TTTGTGCTCC ACGACGGCCT CCGCCGCTAC TTCTCAGAGC CACTGCAGGG TAGGCACAGG 11460  
GCCTGCTGGG GCTCAGGAGC TTGGAGTGTG TGGTTGGGGC AGGCCTGGGG GGTCAATTCTC 11520  
TGGAGCCAGC TGTGTGGCTT CAGGCAGCAG TCAGCGACTT GGCTGCAGTG GGCTGAGAGT 11580  
TCCTTGTCTG AGGAAGGGAG CTGTCATGAG GGAGGGGTCC ATGGCCAGAT GTGAACGCAG 11640  
AATGCACTGA GCCAGGGCCT GGTGACTGCT TGGGAACAGC CTGTGATGAG AAGGGGTTAG 11700  
GCAGCCTTTG CCCCTGGGGC TGCACAGGAA GCCCTAGCCA GCGACCTGGT GACTCCCCCTG 11760  
AGCTGGAAGA GGCTCAGACT CCAGAGGGCA TTGCCTATGG GGCTTTGCAC GGGTGAAGC 11820  
CAGGCCAGCC AAGAGGACCT GTTCCTGCTG GATGTGCTGC ACACCTAGGA ACCTTGTGCT 11880  
TGCCTGCCAC CGCCTCCCTC TGTCCCTTTC TCCATCACAC AGATTTCTGG CATCAGAAAA 11940

ACACCGACTA CAACAACAGC CCTTTCCACA TCTCCCAGTG CTTCGTGCTA AAGCTTAGTG 12000  
 ACCTGCAGTC AGTCAATGTG GGGCTTTTCG ACCAGCATTT CCGGCTGACG GGTGGGTGAC 12060  
 CCTCTGTGCT TTGTCCTATT TCGGGTGAAG GCCAGCATCA CCAGTGGGCT TCCACCTTCC 12120  
 GTACGTGGGT GGGTTATCAT AGACAGTTAT CTCTGTGCTC AAGAGCCACT TCTTACCCGG 12180  
 GGTGGGAGGA AGCAGCTTCA GGAAGTGTG AGAGAGCAGA ACTCACGCTC CAGGGCTCAG 12240  
 AGCAGGAGGT AGGGTGTGCG GCAAGCGCTG GCGGACAG AAGCAGAGTG GGCCCTGGTC 12300  
 TCGGGCAGGA TGTTCCTGAC TCACATTTCC TGAGGAGAGA AAGCTAAGCT CTTTGCTAA 12360  
 TGTCTCTGTC TCCCCCTCCA GAAAAATGCC TCAGCTCTTC CGGCCTGAAG GAATGGCCTC 12420  
 CTCCCGGGCC CCATGATTCT TTCCTGTGTG GGCCCTCCTG GCCCTGGCCT CTGGGCTGAG 12480  
 GCTTGCTAGG GACTCGGGGT GGCTCTAAGG GGCAGGGATA GGGCTGGGGA GCGCCGGCCT 12540  
 GTGGCCCTGA CCAGCCCCTT CTCGTGCAGG TTCCACCCCG ATGCAGGTGG TCACGTGCTT 12600  
 GACGCGGGAC AGCTACCTGA CGCACTGCTT CCTCCAGCAC CTCATGGTCG TGCTGTCCTC 12660  
 TCTGGAACGC ACGCCCTCGC CGGAGCCTGT TGACAAGGAC TTCTACTCCG AGTTTGGGAA 12720  
 CAAGACCACA GGTACCCCTG TCTAGCTCAG GCTGCAGACA GGCTGCCTGG ACAGACGTCA 12780  
 TGGGCCCCAG GGTGGCTCTC TGTGCCCCAG AACCTCTCT GCCTCTATGT CTCTCTTTTC 12840  
 TCACTTAGCT GGCCAGGGTT TTATGTGGGG CTTTTCGATG GCAGAGTCTC CACTCCAGCA 12900  
 GTCCCTCAAC CATCTGGCAG ACACATCTCC AGTGCCCTGCT TTGGGCTCCT GGCCTGTGGG 12960  
 CCCACACTT GGAGCATCCT CTCCTGCCTG TCTCATGCCG GGTCTCTCG GTTGGCTTGG 13020  
 GGCCCTTGGT GCTCCAGCC CCACCAGGG CCGGTTCCAG GCTATAGCCC AGGTGGCATC 13080  
 TCTCTGCAGG GAAGATGGAG AACTACGAGC TGATCCACTC TAGTCGCGTC AAGTTTACCT 13140  
 ACCCAGTGA GGAGGAGATT GGGGACCTGA CGTTCACTGT GGCCAAAAG ATGGCTGAGC 13200  
 CAGAGAAGGC CCCAGCCCTC AGCATCCTGC TGTACGTGCA GGCCTTCCAG GTGGGCATGC 13260  
 CACCCCTGG GTGCTGCAGG GGCCCCCTGC GCCCCAAGAC ACTCCTGCTC ACCAGCTCCG 13320  
 AGATCTTCCT CCTGGATGAG GACTGTGTCC ACTACCCACT GCCCAGGTTT GCCAAAGAGC 13380  
 CGCCGCAGAG AGACAGGTAC CGGCTGGACG ATGGCCGCCG CGTCCGGGAC CTGGACCGAG 13440  
 TGCTCATGGG CTACCAGACC TACCCGAGG CCCTCACCCCT CGTCTTCGAT GACGTGCAAG 13500  
 GTCATGACCT CATGGGCAGT GTCACCCTGG ACCACTTTGG GGAGGTGCCA GGTGGCCCGG 13560  
 CTAGAGCCAG CCAGGGCCGT GAAGTCCAGT GGCAGGTGTT TGTCCCCAGT GCTGAGAGCA 13620  
 GAGAGAAGCT CATCTCGCTG TTGGCTCGCC AGTGGGAGGC CCTGTGTGGC CGTGAGCTGC 13680  
 CTGTCGAGCT CACCGGCTAG CCCAGGCCAC AGCCAGCCTG TCGTGTCCAG CCTGACGCCT 13740  
 ACTGGGGCAG GGCAGCAGGC TTTTGTGTTT TCTAAAAATG TTTTATCCTC CTTTGGGTAC 13800  
 CTTAATTTGA CTGTCCTCGC AGAGAATGTG AACATGTGTG TGTGTTGTGT TAATTCTTTC 13860  
 TCATGTTGGG AGTGAGAATG CCGGGCCCT CAGGGCTGTC GGTGTGCTGT CAGCCTCCCA 13920  
 CAGGTGGTAC AGCCGTGCAC ACCAGTGTCT TGTCTGCTGT TGTGGGACCG TTGTAAACAC 13980



GTGACACTGT GGGTCTGACT TTCTCTTCTA CACGTCCTTT CCTGAAGTGT CGAGTCCAGT 14040  
 CCTTTGTTGC TGTGCTGTT GCTGTTGCTG TTGCTGTTGG CATCTTGCTG CTAATCCTGA 14100  
 GGCTGGTAGC AGAATGCACA TTGGAAGCTC CCACCCCATTA TTGTTCTTCA AAGTGGAGGT 14160  
 CTCCCCTGAT CCAGACAAGT GGGAGAGCCC GTGGGGGCAG GGGACCTGGA GCTGCCAGCA 14220  
 CCAAGCGTGA TTCCTGCTGC CTGTATTCTC TATTCCAATA AAGCAGAGTT TGACACCGTC 14280  
 TGCATCTTCT AAACCAAGGG TCACTGGGAT CGAGTCGACG GCCCTATAGT GAGTCGTATT 14340  
 AGAGCTCGCG GCCGCGAGCT CTAGATGCAT GCTCGAGCGG CCGCCAGTGT GATGGATATC 14400  
 TGCAGAATTC CAGCACACTG GCGGCCGTTA CTAGTGGATC CGAGCTCCAC AGAGGTGGTC 14460  
 GATGAGGGCT GCCCTTTCCC ACATCCTTAG TAGGGGGTTC AAGATGACCC AGACTGTGCC 14520  
 CCTGGGGAGC TTGGAGCCAT GCGGGAGGAT GAGCCATGTG CTGGAGGAGA ACAGGGTAGG 14580  
 ATGGTGTGGG GCTTTTGTAG ACTGTCTAGA AGCAAAGAAG GTCTGCAGTG GAGGTGGTGT 14640  
 CTGAGGTGAA TCTCGAAGGT GAATAGGAGT TGAACGTTAG CAGGCAGAGG GTGGATTGCA 14700  
 GGAGAGCAGC GGCCTGGGCA GGTGCCCAGC GTGGCCCATC AGGGTGCTTC ATGCATGGCT 14760  
 GTGTGCTTGC CATCCTTCCT GCCTGCCTAC CCCCTGCTGC TTCGCTTCAT GGGGGCGTTT 14820  
 GAGCTTGGGC CCACCTGCCT GCCTCGCTTG TGGGCAGAGG ACCCAAGCTG TGTGAGTTGT 14880  
 CCTGTCCCGG GGAGCAGCTG AACTGGTCCG GGGTCTCGAA CTGTGGGGCT CAAAAGGACT 14940  
 CCGGGGTCAT TTCACTGGGG CTGTGCCGAT TCCTGGGGGC TGTNNGAAN GTAAAGGCCT 15000  
 AAAGGGGCTC CCTGGTTANG GCCCTCAANT TTAANAACCT GGGGCCGGGG CCCGGAATTG 15060  
 CCCCCAANTT TGTTTCAACN CCCCTTGGCC TTNGGCNNGG GCAAATTTCC ANGGGGAACC 15120  
 AATGGNTTTC CCCCAAAAAN GGGGCCNTTT TAACCCNTTT CCAAANTTTG GGNCCATAAA 15180  
 AAGGGTGGAN TTCCTGAANG GG 15202

## INFORMATION FOR SEQ ID NO: 22

## SEQUENCE CHARACTERISTICS:

LENGTH: 1070 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

## SEQUENCE DESCRIPTION: SEQ ID NO: 22

Val Cys Leu Asp Asp Thr Val Thr Thr Glu Lys Glu Leu Asp Thr Val  
 1 5 10 15  
 Glu Val Leu Lys Ala Ile Gln Lys Ala Lys Glu Val Lys Ser Lys Leu  
 20 25 30

Ser Asn Pro Glu Lys Lys Gly Gly Glu Asp Ser Arg Leu Ser Ala Ala  
 35 40 45  
 Pro Cys Ile Arg Pro Ser Ser Ser Pro Pro Thr Val Ala Pro Ala Ser  
 50 55 60  
 Ala Ser Leu Pro Gln Pro Ile Leu Ser Asn Gln Gly Ile Met Phe Val  
 65 70 75 80  
 Gln Glu Glu Ala Leu Ala Ser Ser Leu Ser Ser Thr Asp Ser Leu Thr  
 85 90 95  
 Pro Glu His Gln Pro Ile Ala Gln Gly Cys Ser Asp Ser Leu Glu Ser  
 100 105 110  
 Ile Pro Ala Gly Gln Ala Ala Ser Asp Asp Leu Arg Asp Val Pro Gly  
 115 120 125  
 Ala Val Gly Gly Ala Ser Pro Glu His Ala Glu Pro Glu Val Gln Val  
 130 135 140  
 Val Pro Gly Ser Gly Gln Ile Ile Phe Leu Pro Phe Thr Cys Ile Gly  
 145 150 155 160  
 Tyr Thr Ala Thr Asn Gln Asp Phe Ile Gln Arg Leu Ser Thr Leu Ile  
 165 170 175  
 Arg Gln Ala Ile Glu Arg Gln Leu Pro Ala Trp Ile Glu Ala Ala Asn  
 180 185 190  
 Gln Arg Glu Glu Gly Gln Gly Glu Gln Gly Glu Glu Glu Asp Glu Glu  
 195 200 205  
 Glu Glu Glu Glu Glu Asp Val Ala Glu Asn Arg Tyr Phe Glu Met Gly  
 210 215 220  
 Pro Pro Asp Val Glu Glu Glu Glu Gly Gly Gly Gln Gly Glu Glu Glu  
 225 230 235 240  
 Glu Glu Glu Glu Glu Asp Glu Glu Ala Glu Glu Glu Arg Leu Ala Leu  
 245 250 255  
 Glu Trp Ala Leu Gly Ala Asp Glu Asp Phe Leu Leu Glu His Ile Arg  
 260 265 270  
 Ile Leu Lys Val Leu Trp Cys Phe Leu Ile His Val Gln Gly Ser Ile  
 275 280 285  
 Arg Gln Phe Ala Ala Cys Leu Val Leu Thr Asp Phe Gly Ile Ala Val  
 290 295 300

Phe Glu Ile Pro His Gln Glu Ser Arg Gly Ser Ser Gln His Ile Leu  
 305 310 315 320  
 Ser Ser Leu Arg Phe Val Phe Cys Phe Pro His Gly Asp Leu Thr Glu  
 325 330 335  
 Phe Gly Phe Leu Met Pro Glu Leu Cys Leu Val Leu Lys Val Arg His  
 340 345 350  
 Ser Glu Asn Thr Leu Phe Ile Ile Ser Asp Ala Ala Asn Leu His Glu  
 355 360 365  
 Phe His Ala Asp Leu Arg Ser Cys Phe Ala Pro Gln His Met Ala Met  
 370 375 380  
 Leu Cys Ser Pro Ile Leu Tyr Gly Ser His Thr Ser Leu Gln Glu Phe  
 385 390 395 400  
 Leu Arg Gln Leu Leu Thr Phe Tyr Lys Val Ala Gly Gly Cys Gln Glu  
 405 410 415  
 Arg Ser Gln Gly Cys Phe Pro Val Tyr Leu Val Tyr Ser Asp Lys Arg  
 420 425 430  
 Met Val Gln Thr Ala Ala Gly Asp Tyr Ser Gly Asn Ile Glu Trp Ala  
 435 440 445  
 Ser Cys Thr Leu Cys Ser Ala Val Arg Arg Ser Cys Cys Ala Pro Ser  
 450 455 460  
 Glu Ala Val Lys Ser Ala Ala Ile Pro Tyr Trp Leu Leu Leu Thr Pro  
 465 470 475 480  
 Gln His Leu Asn Val Ile Lys Ala Asp Phe Asn Pro Met Pro Asn Arg  
 485 490 495  
 Gly Thr His Asn Cys Arg Asn Arg Asn Ser Phe Lys Leu Ser Arg Val  
 500 505 510  
 Pro Leu Ser Thr Val Leu Leu Asp Pro Thr Arg Ser Cys Thr Gln Pro  
 515 520 525  
 Arg Gly Ala Phe Ala Asp Gly His Val Leu Glu Leu Leu Val Gly Tyr  
 530 535 540  
 Arg Phe Val Thr Ala Ile Phe Val Leu Pro His Glu Lys Phe His Phe  
 545 550 555 560  
 Leu Arg Val Tyr Asn Gln Leu Arg Ala Ser Leu Gln Asp Leu Lys Thr  
 565 570 575

Val Val Ile Ala Lys Thr Pro Gly Thr Gly Gly Ser Pro Gln Gly Ser			
580	585	590	
Phe Ala Asp Gly Gln Pro Ala Glu Arg Arg Ala Ser Asn Asp Gln Arg			
595	600	605	
Pro Gln Glu Val Pro Ala Glu Ala Leu Ala Pro Ala Pro Val Glu Val			
610	615	620	
Pro Ala Pro Ala Pro Ala Ala Ala Ser Ala Ser Gly Pro Ala Lys Thr			
625	630	635	640
Pro Ala Pro Ala Glu Ala Ser Thr Ser Ala Leu Val Pro Glu Glu Thr			
645	650	655	
Pro Val Glu Ala Pro Ala Pro Pro Pro Ala Glu Ala Pro Ala Gln Tyr			
660	665	670	
Pro Ser Glu His Leu Ile Gln Ala Thr Ser Glu Glu Asn Gln Ile Pro			
675	680	685	
Ser His Leu Pro Ala Cys Pro Ser Leu Arg His Val Ala Ser Leu Arg			
690	695	700	
Gly Ser Ala Ile Ile Glu Leu Phe His Ser Ser Ile Ala Glu Val Glu			
705	710	715	720
Asn Glu Glu Leu Arg His Leu Met Trp Ser Ser Val Val Phe Tyr Gln			
725	730	735	
Thr Pro Gly Leu Glu Val Thr Ala Cys Val Leu Leu Ser Thr Lys Ala			
740	745	750	
Val Tyr Phe Val Leu His Asp Gly Leu Arg Arg Tyr Phe Ser Glu Pro			
755	760	765	
Leu Gln Asp Phe Trp His Gln Lys Asn Thr Asp Tyr Asn Asn Ser Pro			
770	775	780	
Phe His Ile Ser Gln Cys Phe Val Leu Lys Leu Ser Asp Leu Gln Ser			
785	790	795	800
Val Asn Val Gly Leu Phe Asp Gln His Phe Arg Leu Thr Gly Ser Thr			
805	810	815	
Pro Met Gln Val Val Thr Cys Leu Thr Arg Asp Ser Tyr Leu Thr His			
820	825	830	
Cys Phe Leu Gln His Leu Met Val Val Leu Ser Ser Leu Glu Arg Thr			
835	840	845	

Pro Ser Pro Glu Pro Val Asp Lys Asp Phe Tyr Ser Glu Phe Gly Asn  
 850 855 860  
 Lys Thr Thr Gly Lys Met Glu Asn Tyr Glu Leu Ile His Ser Ser Arg  
 865 870 875 880  
 Val Lys Phe Thr Tyr Pro Ser Glu Glu Glu Ile Gly Asp Leu Thr Phe  
 885 890 895  
 Thr Val Ala Gln Lys Met Ala Glu Pro Glu Lys Ala Pro Ala Leu Ser  
 900 905 910  
 Ile Leu Leu Tyr Val Gln Ala Phe Gln Val Gly Met Pro Pro Pro Gly  
 915 920 925  
 Cys Cys Arg Gly Pro Leu Arg Pro Lys Thr Leu Leu Leu Thr Ser Ser  
 930 935 940  
 Glu Ile Phe Leu Leu Asp Glu Asp Cys Val His Tyr Pro Leu Pro Glu  
 945 950 955 960  
 Phe Ala Lys Glu Pro Pro Gln Arg Asp Arg Tyr Arg Leu Asp Asp Gly  
 965 970 975  
 Arg Arg Val Arg Asp Leu Asp Arg Val Leu Met Gly Tyr Gln Thr Tyr  
 980 985 990  
 Pro Gln Ala Leu Thr Leu Val Phe Asp Asp Val Gln Gly His Asp Leu  
 995 1000 1005  
 Met Gly Ser Val Thr Leu Asp His Phe Gly Glu Val Pro Gly Gly Pro  
 1010 1015 1020  
 Ala Arg Ala Ser Gln Gly Arg Glu Val Gln Trp Gln Val Phe Val Pro  
 1025 1030 1035 1040  
 Ser Ala Glu Ser Arg Glu Lys Leu Ile Ser Leu Leu Ala Arg Gln Trp  
 1045 1050 1055  
 Glu Ala Leu Cys Gly Arg Glu Leu Pro Val Glu Leu Thr Gly  
 1060 1065 1070

WHAT IS CLAIMED IS:

CLAIMS

1. A DNA molecule encoding for a polypeptide including an amino acid sequence which is receptive to imidazoline compounds, said DNA molecule containing a DNA sequence with at least 75% sequence similarity with the DNA sequence shown in SEQ ID No. 4.
2. A DNA molecule according to claim 1, containing a DNA sequence with at least 75% sequence similarity with the DNA sequence shown in SEQ ID No. 2.
3. A DNA molecule according to claim 2, containing a DNA sequence with at least 75% sequence similarity with the DNA sequence of SEQ ID No. 3.
4. A DNA molecule according to claim 3, containing a DNA sequence with at least 75% sequence similarity with the DNA sequence of SEQ ID No. 1.
5. A DNA molecule according to any one of claims 1 to 4, containing a DNA sequence with at least 80% sequence similarity with the sequence of said SEQ ID No.
6. A DNA molecule according to any one of claims 1 to 4, containing a DNA sequence with at least 85% sequence similarity with the sequence of said SEQ ID No.

7. A DNA molecule according to any one of claims 1 to 4, containing a DNA sequence with at least 90% sequence similarity with the sequence of said SEQ ID No.
8. A DNA molecule according to any one of claims 1 to 4, containing a DNA sequence with at least 95% sequence similarity with the sequence of said SEQ ID No.
9. A DNA molecule according to claim 1, which is deposited with the ATCC under deposit accession no. ATCC 209217.
10. A genomic DNA molecule encoding for a polypeptide including an amino acid sequence which is receptive to imidazoline compounds, and wherein exon portions of said genomic DNA molecule include the DNA sequence as defined in claim 1.
11. A genomic DNA molecule according to claim 10, which is deposited with the ATCC under deposit accession no. ATCC 209216.
12. A 1110 bp ApaI-EcoRI restriction fragment of the DNA molecule according to claim 1.
13. A 1.85 kb EcoRI restriction fragment of the DNA molecule according to claim 4.

14. A vector containing a DNA sequence as defined in any one of claims 1-13.

15. A host cell transfected with a vector as defined in claim 14.

16. An isolated polypeptide including a site which is receptive to imidazoline compounds, said polypeptide containing an amino acid sequence with at least 80% sequence similarity with the amino acid sequence shown in SEQ ID No. 6.

17. A polypeptide as defined in claim 16, having a molecular weight of about 35 to 45 kDa.

18. A polypeptide as defined in claim 17, having a molecular weight of about 37 kDa.

19. An isolated polypeptide including a site which is receptive to imidazoline compounds, said polypeptide containing an amino acid sequence with at least 80% sequence similarity with the amino acid sequence shown in SEQ ID No. 5.

20. A polypeptide as defined in claim 19, having a molecular weight of about 60 to 85 kDa.



21. A polypeptide as defined in claim 20, having a molecular weight of about 70 kDa.
22. A fragment of the amino acid sequence shown in SEQ ID No. 5 or 6, which fragment is receptive to imidazoline compounds.
23. A polypeptide according to any one of claims 16 to 22, which is immunoreactive with at least one of Reis antiserum and Dontenwill antiserum.
24. A polypeptide according to any one of claims 16 to 23, which is a human polypeptide.
25. A method of producing an isolated polypeptide including an amino acid sequence which is receptive to imidazoline compounds, said method comprising:
  - transfecting a host cell with a vector as defined in claim 14; and
  - culturing the transfected host cell in a culture medium to express the polypeptide.
26. An isolated polypeptide including an amino acid sequence which is receptive to imidazoline compounds, which polypeptide is expressed by the method of claim 25.

27. A method of screening for a ligand of an imidazoline receptor, which method comprises:

culturing a host cell as defined in claim 15 in a culture medium to express a polypeptide including an amino acid sequence which is receptive to imidazoline compounds;

contacting said polypeptide with a labelled ligand for the imidazoline receptor under conditions effective to bind the labelled ligand thereto;

contacting said polypeptide with a candidate ligand; and

detecting any displacement of the labelled ligand from said polypeptide, wherein displacement signifies that the candidate ligand is a ligand for the imidazoline receptor.

28. The method of claim 27, wherein said contacting steps are performed in an intact cultured host cell.

29. The method of claim 27, further comprising isolating the cell membrane of said cultured host cell prior to performing said contacting steps.

30. The method of claim 27, wherein said contacting of said imidazoline receptive polypeptide with said candidate ligand is conducted at a plurality of candidate ligand concentrations.

31. The method of claim 27, wherein the labelled ligand is radiolabelled.

32. A method of obtaining a DNA material encoding a polypeptide which is receptive to imidazoline compounds, said method comprising:

providing a labelled DNA probe by labelling a DNA molecule identical or complementary to a DNA molecule as defined in any one of claims 1 to 9 or a restriction fragment thereof;

contacting said DNA probe with genetic material suspected of encoding said imidazoline receptive polypeptide;

hybridizing said DNA probe and said genetic material under stringent hybridization conditions;

identifying any portion of the genetic material which hybridizes to said DNA probe; and

isolating said identified material.

33. A method according to claim 32, wherein the genetic material is derived from a library selected from the group consisting of RNA library, cDNA library and genomic DNA library.

34. A method according to claim 33, wherein said library is a human library.

35. A method according to claim 32, wherein the labelled DNA probe is provided by labelling a restriction fragment according to claim 12 or 13.

36. A method of raising antibodies immunoreactive with a polypeptide which is receptive to an imidazoline compound, which method comprises:

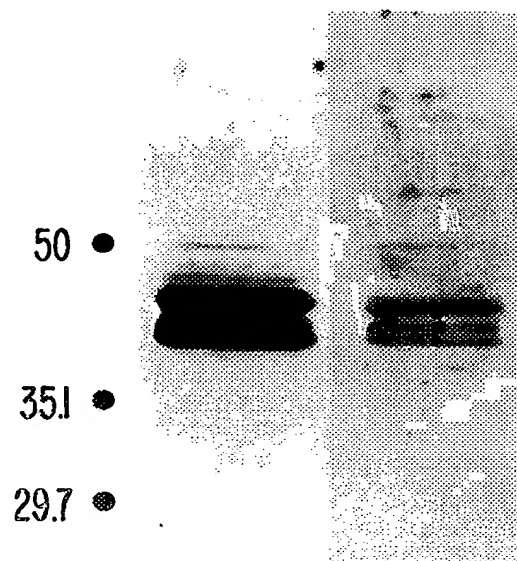
injecting an animal with a polypeptide as defined in any one of claims 16 to 24 and 26; and

isolating antibodies produced by the animal.

**FIG. 1A**

HIPPOCAMPUS

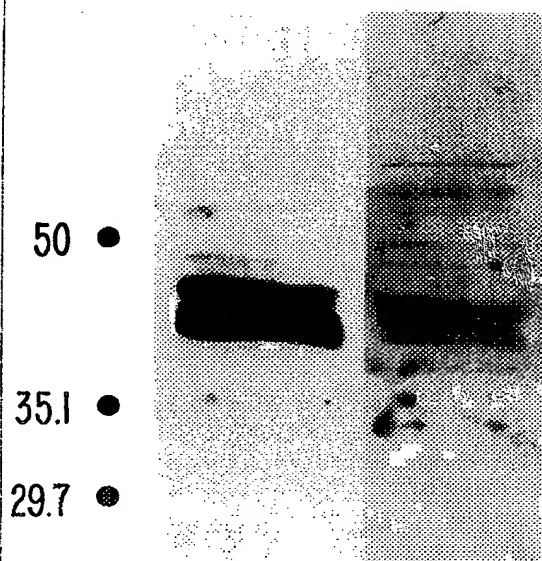
**1 2**



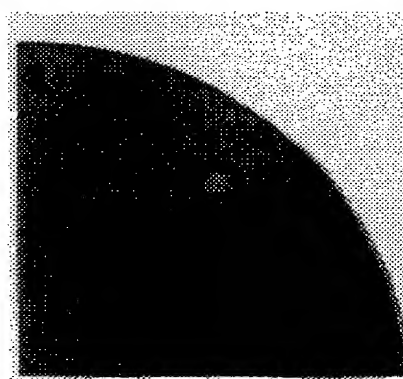
**FIG. 1B**

NRL

**1 2**

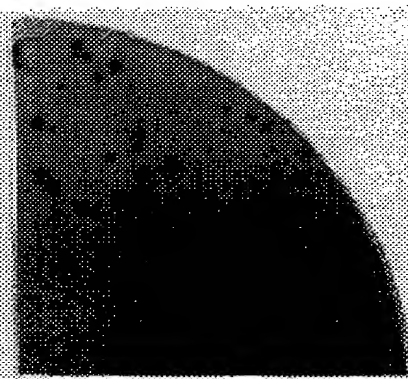


**FIG. 2A**



REIS AB  
1:15,000 DILUTION

**FIG. 2B**

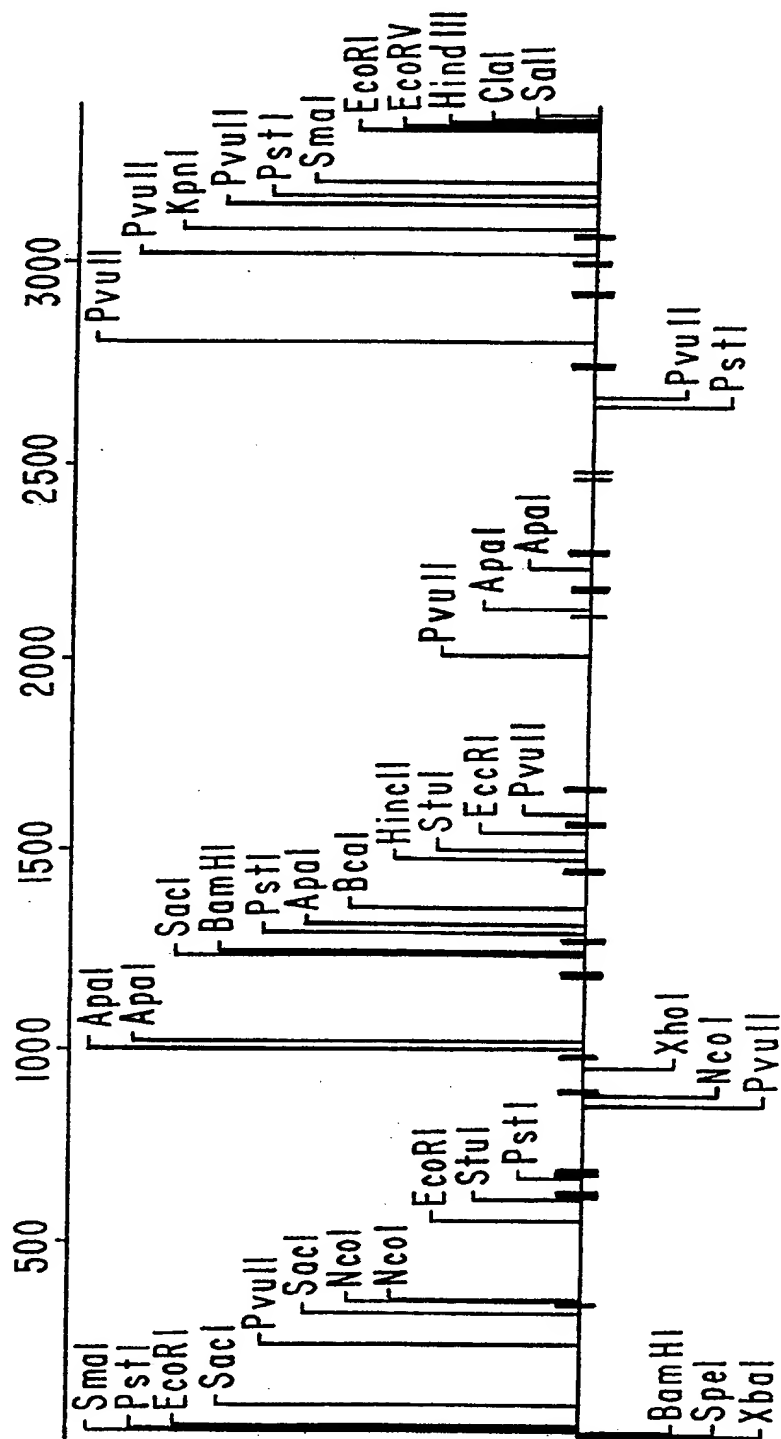


DONTWILL AB  
1:20,000 DILUTION



2/6

FIG. 3

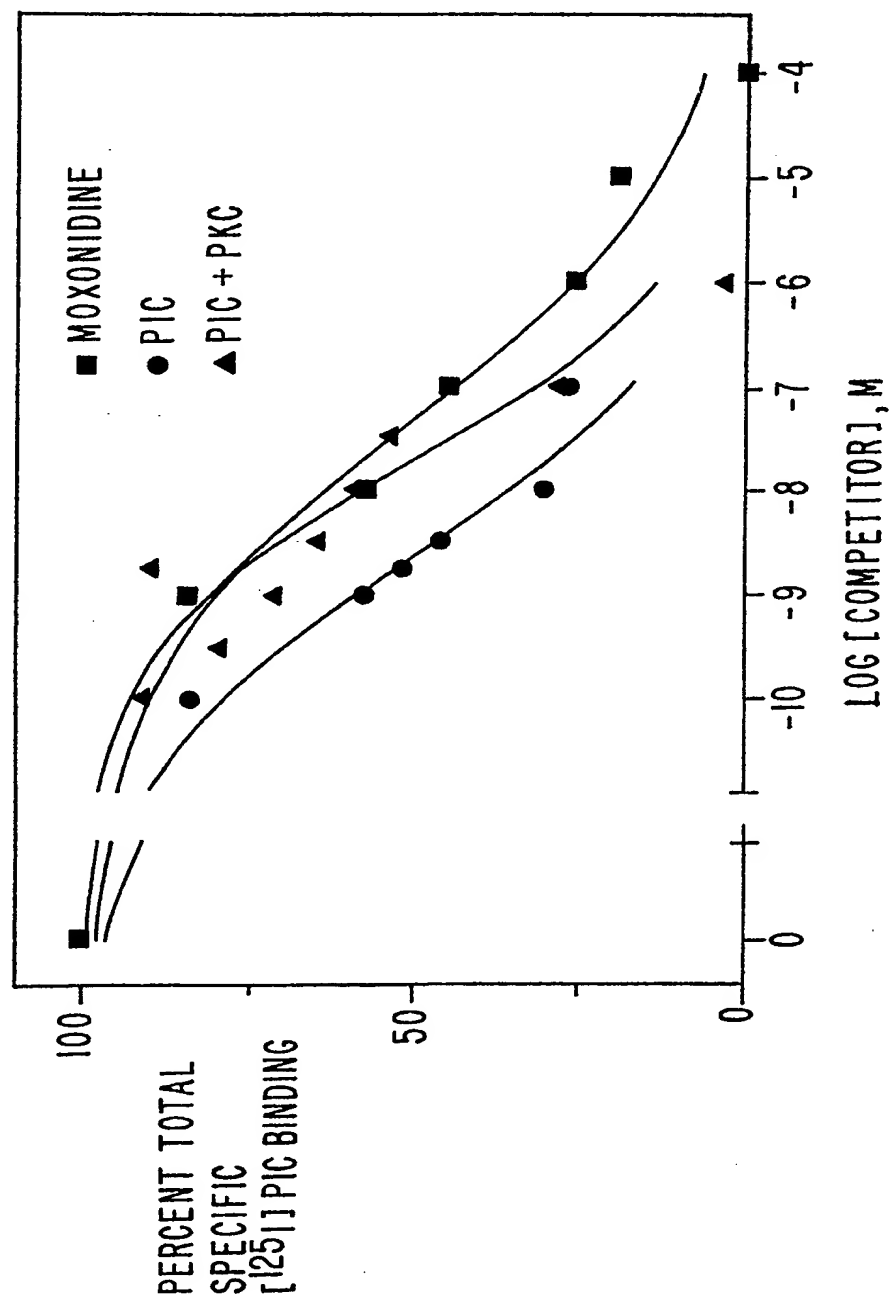






3/6

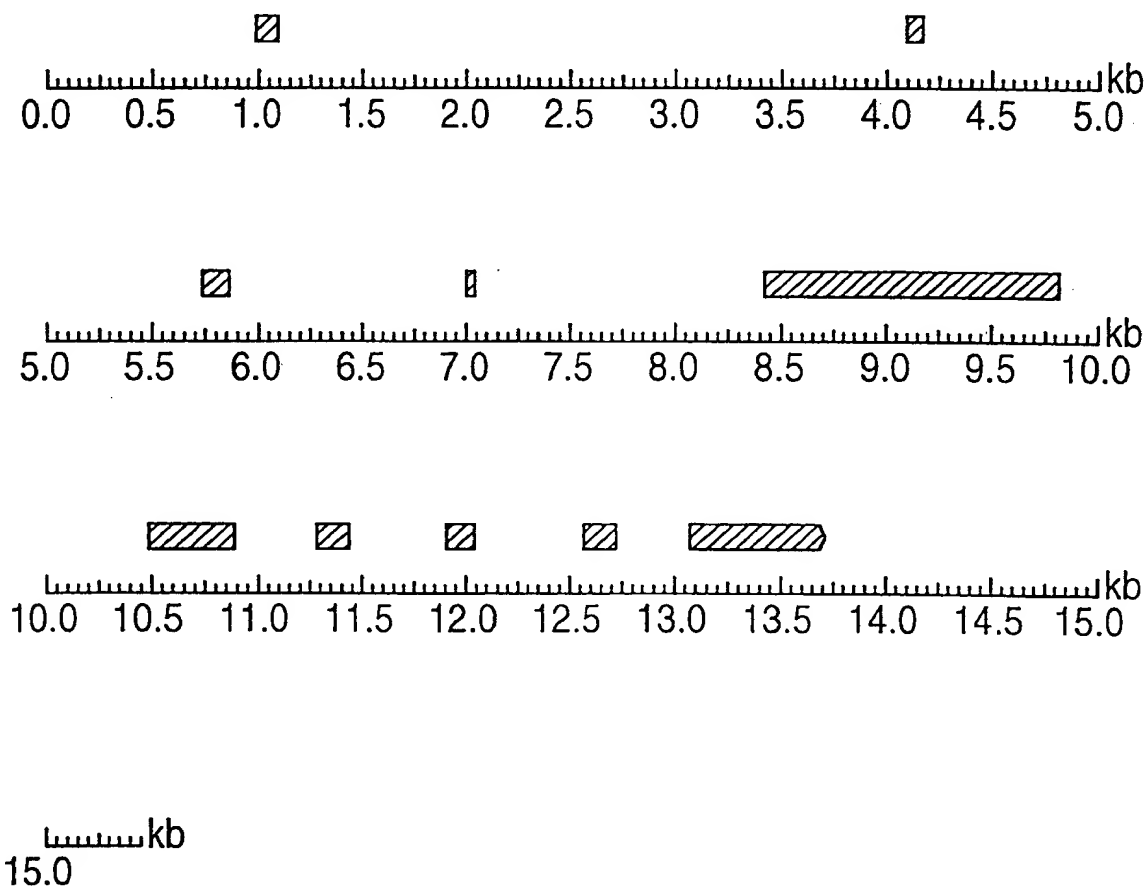
FIG. 4





4/6

FIG. 5

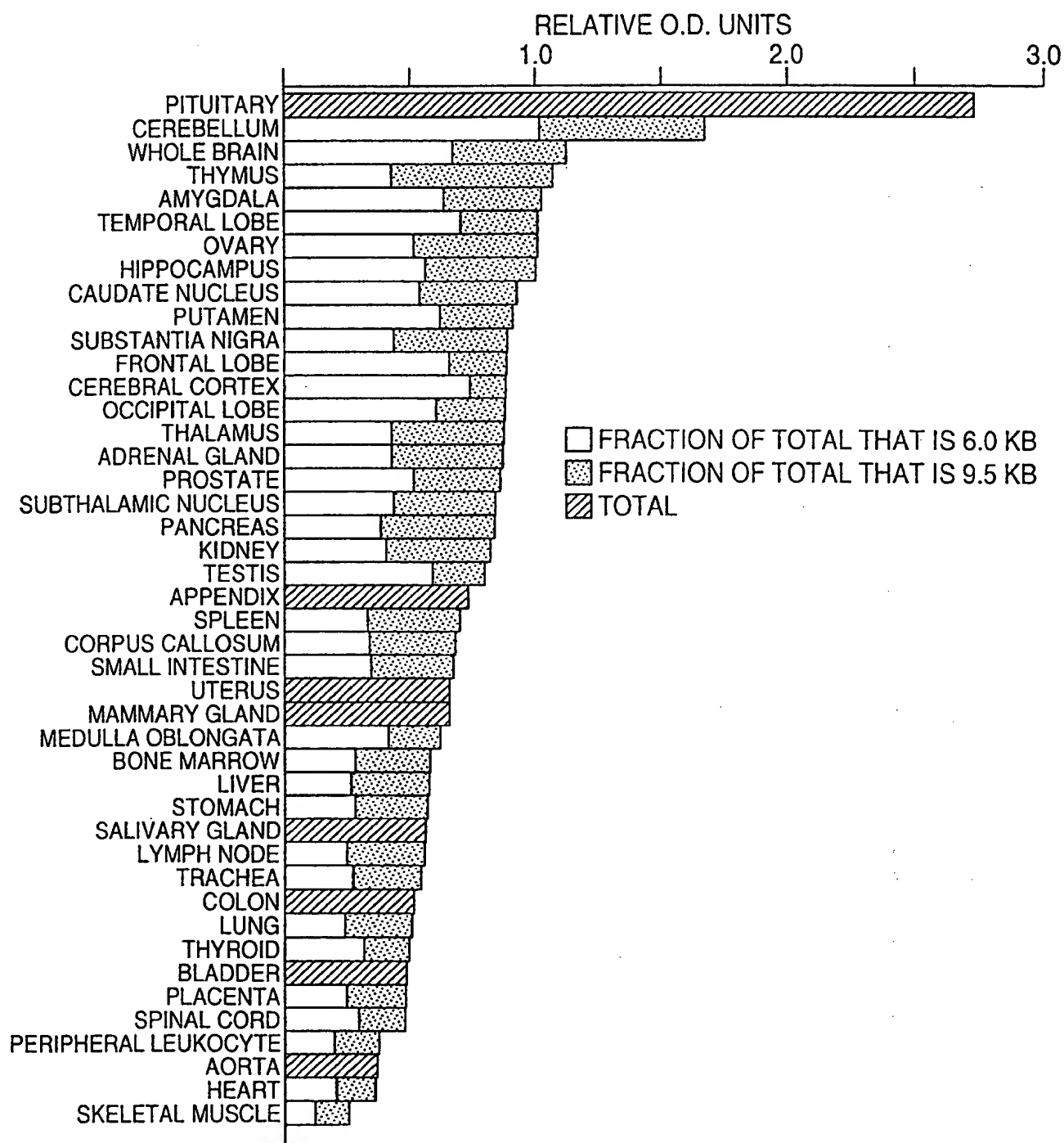


KEY: INITIAL EXON INTERNAL EXON TERMINAL EXON SINGLE-EXON GENE OPTIMAL EXON SUBOPTIMAL EXON



5/6

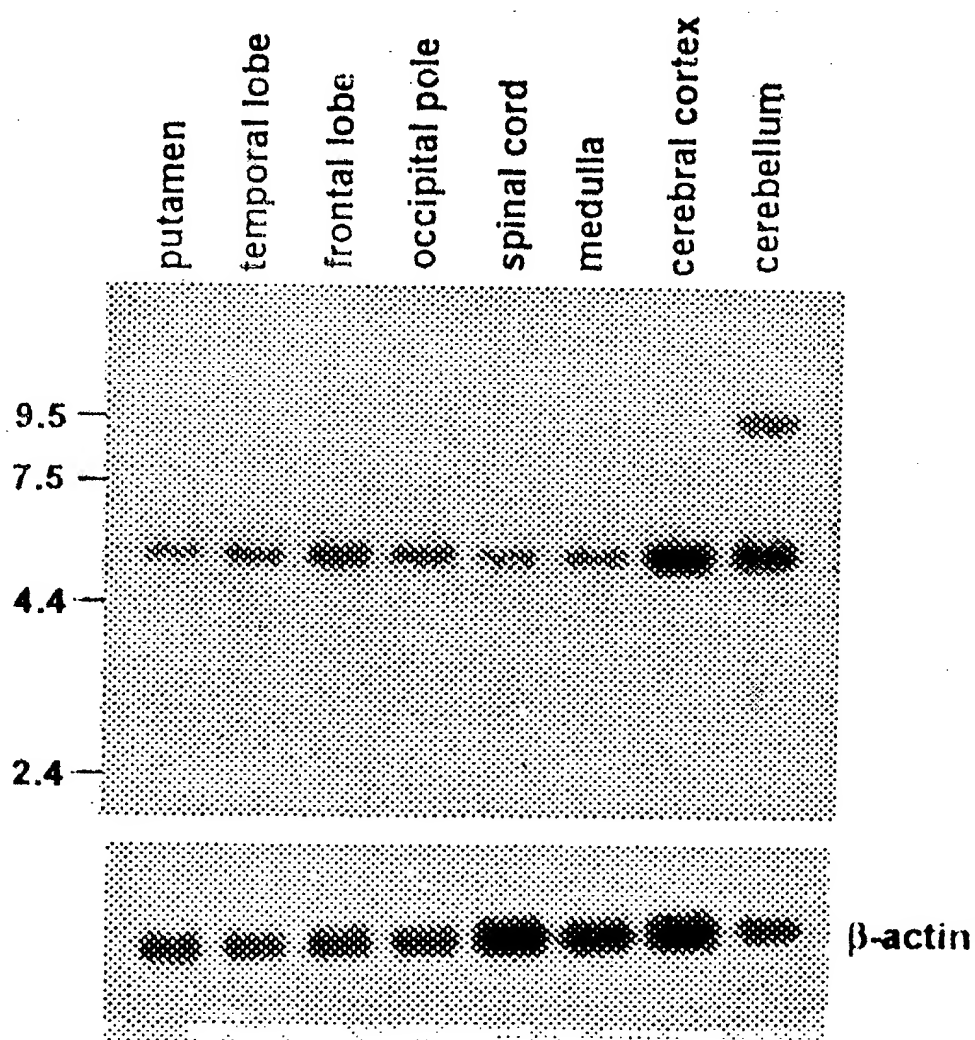
FIG. 6A





6/6

FIG. 6B







# INTERNATIONAL SEARCH REPORT

Int lional Application No  
PCT/US 97/15695

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C07K14/705 A61K38/17 C12N5/10 C12N15/12

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL Emest5:Hs1228705 Accession number AA428250, 25 May 1997 "WashU-Merck EST Project 1997" XP002064993	19, 24
Y	Unpublished see abstract	20-22
X	DATABASE EMBL Emest5:Hs1228705 Accession number AA428250, 25 May 1997 "WashU-Merck EST Project 1997" XP002064994 Unpublished see abstract	2, 5-8

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

14 May 1998

Date of mailing of the international search report

15/06/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Halle, F

# INTERNATIONAL SEARCH REPORT

Int lional Application No

PCT/US 97/15695

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL  Emest4:Hs1190779  Accession number AA287493, 12 April 1997  "National Cancer Institute, Cancer Genome  Anatomy Project (CGAP), Tumor Gene Index"  XP002064995  Unpublished  see abstract</p> <p style="text-align: center;">---</p>	4-8
X	<p>DATABASE EMBL  Emest4:Hs1190779  Accession number AA287493, 12 April 1997  "National Cancer Institute, Cancer Genome  Anatomy Project (CGAP), Tumor Gene Index"  XP002064996  Unpublished  see abstract</p> <p style="text-align: center;">---</p>	3,5-8
X	<p>DATABASE EMBL  Emest13:Mmw3021  Accession number W91302, 9 July 1996  "The WashU-HHMI Mouse EST Project"  XP002064997  see abstract</p> <p style="text-align: center;">---</p>	3,5-7
X	<p>DATABASE EMBL  R54u006:Rs782  Accession number H31782, 30 September 1995  "Comparative expressed-sequence-tag  analysis of differential gene expression  profiles in PC-12 cells before and after  nerve growth factor treatment"  XP002064998  Unpublished  see abstract  &amp; Proc. Natl. Acad. Sci. USA 92:8303-8307  (1995)</p> <p style="text-align: center;">---</p>	2,5
X	<p>DATABASE EMBL  Emest10:Hsc1sa031  Accession number F06998, 15 February 1995  "The Genexpress cDNA program"  XP002064999</p>	19,24
Y	<p>Unpublished  see abstract</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	20-22

# INTERNATIONAL SEARCH REPORT

Int. l. Application No

PCT/US 97/15695

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL Emest:Hs1442 Accession number T06144, 2 September 1993 "3,400 new expressed sequence tags identify diversity of transcripts in human brain" XP002065000	16,24
Y	see abstract  & Nat. Genet. 4:256-267 (1993) ----	10,12, 14,15, 17,18,22
X	DATABASE EMBL Emest5:Hs1442 Accession number T06144, 2 September 1993 "3,400 new expressed sequence tags identify diversity of transcripts in human brain" XP002065001	1,5-7
Y	see abstract  & Nat. Genet. 4:256-267 (1993) ----	10,12, 14,15, 17,18
E	WO 97 31945 A (UNIV MISSISSIPPI MEDICAL CENTE, 04.09.1997) 4 September 1997 see the whole document -----	1-36

# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/US 97/15695

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9731945 A	04-09-97	NONE	